CUK-H05918-2-KP128

THE JOURNAL OF PHILOSOPHY

②

VOLUME LXXXVIII, NUMBER 1
JANUARY 1991

page
5 *A Modal Argument for Narrow Content* Jerry A. Fodor
27 *Real Patterns* Daniel C. Dennett

BOOK REVIEWS
52 Arthur W. Collins: *The Nature of Mental Things*
Anita Avramides

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1991 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER Second class postage paid at New York, NY, and additional mailing offices Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

Calcutta University Library
ALIPORE CAMPUS

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII
JANUARY–DECEMBER 1991

Edited by
Bernard Berofsky, Arthur C. Danto
Kent Greenawalt, Charles Larmore
Isaac Levi, Sidney Morgenbesser
Mary Mothersill

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY was founded in 1904 as THE JOURNAL OF PHILOSOPHY, PSYCHOLOGY AND SCIENTIFIC METHODS by Frederick J. E. Woodbridge and J. McKeen Cattell. In 1906 Wendell T. Bush became associated with the JOURNAL as co-editor. In 1923, the JOURNAL was incorporated in the State of New York under its present name. Since its founding it has been published from Columbia University and printed by the Lancaster Press in Lancaster, Pennsylvania.

K-P-128

Calcutta University Library

THE JOURNAL OF PHILOSOPHY

STATEMENT OF COPYRIGHT POLICY

We would like to draw your attention to the protection copyright affords against multiple photocopying by libraries, instructors, and copy centers. The 1976 copyright law provides that photocopies in excess of "fair use" copying cannot be made without authorization by the copyright owner—"fair use" meaning single copying for personal use. Accordingly, the *Journal* authorizes personal or educational multiple copying only if a request is made in writing, and the fee of $.05 per page/per copy is paid directly to us. *Each copy must include a notice of copyright.* We will turn over to the authors any substantial payments (over $20.00, an administrative fee) obtained from all photocopying. Inquiries about photocopying may be sent to the Managing Editor.

© Copyright 1991 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices
Send address changes to *Journal of Philosophy*, 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NO. 1, JANUARY 1991

A MODAL ARGUMENT FOR NARROW CONTENT*

Here is a modern antinomy. On the one hand, there is argument *A*:

Argument *A*:

1. My twin and I are molecular duplicates.
2. Therefore our (actual and counterfactual) behaviors are identical in relevant respects.
3. Therefore the causal powers of our mental states are identical in relevant respects.
4. Therefore my twin and I belong to the same natural kind for purposes of psychological explanation and "individualism" is true.

But, on the other hand, there is argument *B*:

Argument *B*:

1′ My twin and I are molecular duplicates.
2′ Nevertheless, our (actual and counterfactual) behaviors are different in relevant respects.
3′ Therefore the causal powers of our mental states are different in relevant respects.
4′ Therefore my twin and I belong to different natural kinds for purposes of psychological explanation and "individualism" is false.

At least one of these arguments must be unsound. Which one? And what is wrong with it?

In chapter 2 of *Psychosemantics*,[1] I offered some considerations intended to advance the cause of argument *A*. They were supposed

---

* This paper is deeply indebted to conversations with Georges Rey and Steve Stich. Help from Ned Block, Anne Jacobson, Tim Maudlin, Colin McGinn, Brian McLaughlin, and Stephen Schiffer is also gratefully acknowledged, as are edifying emails from Fred Adams, Joe Levine, and David Rosenthal.

[1] Cambridge: MIT, 1987.

0022-362X/91/8801/5–26

to show that mental states that differ only in "broad" intentional properties (the sorts of intentional properties that the mental states of molecular twins may fail to share) ipso facto do not differ in causal powers; hence that mere differences in broad intentional content do not determine differences in natural kinds for purposes of psychological explanation. The arguments I offered in *Psychosemantics* were not, however, greeted with unequivocal enthusiasm. Such was their subtlety, in fact, that I am not, myself, always quite sure how they were supposed to go. In the present paper, I propose to have another try at bolstering argument *A*. I shall use many of the same materials that I did in *Psychosemantics,* but I shall put the pieces together somewhat differently. In passing I shall have a word or two to say about some of the comments that the *Psychosemantics* arguments have provoked.

A preliminary remark, however, before we are under way. For most philosophical purposes, it may not really matter much how the individualism issue turns out. For example, I do not think that a resolution in favor of argument *A* (that is, in favor of individualism) would affect the status of "externalism" in semantics. Externalism is independent of individualism because, whatever the *explanatory* status of broad content, it is not in dispute (anyhow, it is not in *this* dispute) that the content of my twin's *water* thoughts differs from the content of mine; or that 'water' means something different in my mouth and in his; or that these semantical differences derive from differences in our respective head/world relations. (Presumably they derive from the fact that, whereas the causal history of my *water* thoughts connects them to samples of H_2O, the causal history of his connects them to samples of *XYZ*). If, in short, the intuitions about twins make a case for content externalism at all, then that case stands whether or not broad intentional states determine natural kinds for the purposes of causal explanation in psychology.

Similarly—and pace, for example, Steven Stich[2]—it is by no means obvious that the scientific vindication of intentional realism depends on how the issues about individualism are resolved. Suppose that there is some metaphysical argument to show that causal explanation in psychology requires the individuation of mental states to be individualistic (so that the mental states of molecular twins are ipso facto type-identical, as in argument *A*.) It would follow that psychological explanation is not, strictly speaking, a species of belief/desire explanation since it is common ground that beliefs and desires are individuated broadly. But it would still be open whether psychological explanations are species of *intentional* explanations. That would depend on whether an individualistic notion of "intentional state" suitable for the purposes of psychological explanation can be constructed. (For more on this, see *Psychosemantics*.)

---

[2] *From Folk Psychology to Cognitive Science* (Cambridge: MIT, 1983)

In short, you can do quite a lot of business in semantics and the philosophy of mind while benignly neglecting the issues about individualism. Still, I think it is worth trying to get these issues straight. We are about to see that they raise some interesting questions about the family of notions that include causal explanation, causal power, type identity, and natural kind. Getting some of this stuff sorted out may therefore be of use to metaphysics and the philosophy of science, even if it leaves things more or less unaltered in semantics and the philosophy of mind. So here we go.

Argument *A* says that, in virtue of our molecular identity, my twin's behavior and my behavior are identical *in all relevant respects*. Not, of course, that they are identical *tout court*. On the contrary, just as it is common ground in this discussion that the *mental states* of twins can differ in certain of their intentional properties, so it is common ground that twin *behaviors* can differ under some of their intentional descriptions. Indeed, it may be that the second concession is entailed by the first, since it is plausible that the intentional properties of behaviors are inherited from the intentional contents of their mental causes. Thus, it is plausibly because I can think about water and my twin cannot that I but not my twin can reach for water, drill for water, or recommend water to a thirsty friend. A committed individualist might wish to argue that none of these is really, strictu dictu, an example of *behavioral* description; but that is not a tack that I propose to take.

So then: twins can differ in the contents of their mental states; and, in consequence of these mental state differences, the behaviors of twins can differ in certain of their intentional properties. I shall take for granted (what is not, however, uncontested) that these differences in intentional properties are the *only* (relevant) differences between the twins' behaviors. That is, I am taking it for granted that the behaviors of twins are ipso facto identical in all nonintentional properties that are relevant to psychological taxonomy. That is, of course, stronger than assuming what, I think, literally everybody agrees about, namely, that our behaviors are identical under *physical* descriptions, i.e., identical qua motions.

Here is an apparent exception to the stronger claim: if I utter 'Gimme water', then, if all goes well, I get water; but if my twin utters 'Gimme water', then, if all goes well, he gets twater. So maybe there are nonintentional descriptions under which our behaviors are relevantly different after all. Second thoughts, however, make this difference go away; it is an artifact of the boring consideration that, whereas my utterance happened on earth, his happened on twin-earth. The salient consideration is that, if twin-me had uttered 'Gimme water' here, he would have gotten water; and if I had uttered 'Gimme water' on twin-earth, I would have gotten twater. To put it slightly differently, my 'Gimme water' does not get water *come what may;* at best it gets water *in certain circumstances*. But in *those* circumstances, his 'Gimme water' gets water, too.[3]

The moral is that you have to judge identity and difference of causal powers in a way that bears the counterfactuals in mind, namely, *across* contexts rather than *within* contexts.[4] That does not get us out of the woods, however. For notice that, in the case where I utter 'Gimme water' on $Earth_2$ and my twin utters it here, neither of us gets what he asks for. That is, whatever the context of utterance, my utterance is a water request and his utterance is a twater request. So our behaviors remain relevantly different under these intentional descriptions *even by the across-context test*. It is this residual difference between the behaviors—their cross-context difference under certain intentional descriptions—which is the challenge to individualism and local supervenience.

---

[3] Here is a related case—suggested by an example of Colin McGinn's—that I take to be susceptible to the same sort of treatment. Suppose on $earth_2$, there is not only twin-water, but also twin-salt (it is LCaN rather than NaCl). And suppose the fact is that thinking about salt makes you want water (whereas thinking about LCaN makes your twin want *XYZ*). So, would there not then be a difference between the causal powers of salt thoughts and twin-salt thoughts in virtue of these differences between the contents of the wants that they cause? No, because salt thoughts do not have the power to cause water wants *come what may;* what they have is the power to cause water wants *in somebody who has the concept water*. (That is, in somebody who has the appropriate causal/historical—or whatever—connections to H_2O.) But, of course, twin-salt thoughts have *that* power too (just as salt thoughts have the power to cause twater wants in somebody who has the appropriate causal/historical connections to *XYZ*.) Here as elsewhere, one applies the cross-context test by asking whether *A* would have the same effects as *B* does have if *A* were to interact with the same things (in the present case, with the same *mental* things) with which *B* does interact.

The box score is: the cross-context test shows that *certain differences that their effects have under intentional description* (namely, causing NaCl wants versus causing LCaN wants) and *certain differences that their effects have under nonintentional description* (see text) do *not* make the difference between having water thoughts and having twater thoughts a difference of causal power. We are about to see, however, that there are differences between the effects that twin mental states exhibit under intentional description which *survive* the cross-context test. These are the ones that, prima facie, make trouble for individualism.

[4] This, surely, is the intuitively natural way to compare causal powers. Consider:

—"Cats raised in Manhattan unable to climb trees," top scientist says.
—Why, that's rather disturbing. How do you explain it?
—There aren't any trees in Manhattan for them to climb.
—Oh.

(For more on this, see the Appendix)

So then, the question about individualism is: Do the twins' mental states belong to different natural kinds (do they have different causal powers) in virtue of differences in the intentional properties of the twins' behavior for which they are responsible? Since 'causal power', 'natural kind', and the like are, of course, technical terms, this question is not possessed of the highest degree of clarity. We shall see, however, that there are some clear intuitions about cases, and they will do for the purposes at hand.

It should be evident that the kind of question we are raising about the intentional states of twins can also arise in cases that have nothing in particular to do with intentionality. Suppose we have a pair of causes *C*1, *C*2, together with their respective effects *E*1, *E*2. Assume that:

> *C*1 differs from *C*2 in that *C*1 has cause property *CP*1 where *C*2 has cause property *CP*2.
>
> *E*1 differs from *E*2 in that *E*1 has effect property *EP*1 and *E*2 has effect property *EP*2.
>
> The difference between *C*1 and *C*2 is responsible for the difference between *E*1 and *E*2 in the sense that, if *C*1 had had *CP*2 rather than *CP*1, then *E*1 would have *EP*2 rather than *EP*1; and if *C*2 had had *CP*1 rather than *CP*2, *E*2 would have had *EP*1 rather than *EP*2.[5]

Call this *schema S*.[6] And now, what we want to know is: Which instances of schema *S* are cases where the difference between having *CP*1 and having *CP*2 is a difference in causal power in virtue of its responsibility for the difference between *E*1 and *E*2? (I shall often abbreviate this to 'When is having *CP*1 rather than *CP*2 a causal power?') If we knew the answer to this general question, then we would know whether, in particular, the difference between having water thoughts and having twater thoughts is a difference in causal power in virtue of its being responsible for the difference between my producing water behaviors and my twin's producing twater behaviors.[7] And, if we knew *that* we would know whether individualism is true; which is what we started out wanting to know, as the patient reader may recall.

---

[5] It is important, in order that relevant questions not be begged, that this is *all* that is required for the fact that *C*1 has *CP*1 to be "responsible for" the fact that *E*1 has *EP*1 (and similarly, mutatis mutandis, for the fact that *C*2 has *CP*2 to be responsible for the fact that *E*2 has *EP*2.) In like spirit, nothing in the examples will depend on stressing the requirement that the *C*s and the *E*s be related as causes and effects, so long as it is assumed that the difference between the *C*s is responsible for the difference between *E*s in the sense just specified.

[6] It will ease the exposition if we think of schema *S* sometimes as relating events and sometimes as relating event types. I shall exploit this ambiguity in what follows; but nothing in the argument turns on it.

Now, a plausible first reaction to this is that it is just not a philosophical issue. For, one might say, the question being raised is when the fact that a generalization supports counterfactuals shows that it has explanatory status. And the answer will depend, in the usual way, on systematic questions about the simplicity, plausibility, power, and so forth of the explanations, and on whether alternative counterfactual supporting generalizations are available. Some of the things that Tyler Burge[8] says about the desirability of not apriorizing about the taxonomy that psychological explanation requires suggest that he would approve of this line of thought; for example:

> It is a mistake . . . to allow ontological preconceptions that have only philosophical underpinning to affect one's interpretation of scientific enterprises. It is a larger mistake to allow them to dictate the sorts of explanatory kinds that are deemed admissible for explanation (*ibid.*).

But though I would not have thought that apriorism is among my major methodological vices, still this does strike me as perhaps a little prim. It would be surprising if we could give interesting a priori *sufficient* conditions for when a difference in the properties of causes constitutes a difference in their causal powers. But it seems likely that we can give a priori arguments for some *necessary* conditions. After all, the commitment to causal explanation presumably has *some* methodological consequences qua commitment to causal explanation; and it ought to be possible to tease these out by reflecting on what kinds of things causal explanations are.

And, in fact, it does seem reasonably clear, a priori, that some instances of schema *S* are not bona fide. Consider, for a thoroughly trivial example, the case where *EP*1 is the property of being the effect of *C*1, and *EP*2 is the property of being the effect of *C*2. There is, of course, a property of *C*1 in virtue of which its effects are effects of *C*1, namely, the property of being *C*1. Correspondingly, there is a property of *C*2 in virtue of which its effects are effects of *C*2, namely, the property of being *C*2. Counterfactual support goes through in the required way; if the cause of *E*1 had had the property of being *C*2 rather than *C*1, then *E*1 would have had the property *EP*2 rather than the property $EP1$. But it seems a priori obvious that this is not a case where having $CP1$ is a causal power of C's in virtue of its responsibility for E's having $EP1$. One of the properties of my effects that your effects cannot have, however hard you try, is *the property of being caused by me*. But I take it to be simply obvious that this difference in our effects does not make the property of *being me rather than you* a causal power. I am not a unit natural kind in virtue of my unique power to cause effects that are effects of me.

---

[7] Reminder: 'Water behavior' means not *behavior that has to do with water* but *behavior that has a reference to water in its intentional description*. We are assuming (see above) that the only relevant descriptions under which the behaviors of twins differ are intentional.

[8] "Individuation and Causation in Psychology" (manuscript, UCLA, 1989).

It seems clear a priori, then, that not every case in which a difference between causes is responsible for a difference in effects is a case where the difference in causes is a difference in their causal powers. There are, in fact, plenty of examples. I can define a property *being an H-particle* that is satisfied by any x at time t iff [(x is a physical particle at t) & (the coin in my hand at t is heads up)]. And correspondingly for T-particles. So a difference between properties of coins in my hand (namely, the difference between being heads up and being tails up) is responsible for the difference between the state of affairs in which all the particles in the universe are H-particles and the state of affairs in which all the particles in the universe are T-particles. But, of course, the difference between being heads up and being tails up does not count as a causal power in virtue of its responsibility for this difference in the particles.

Or again: in virtue of my having siblings, I am able to have sons who are nephews. A molecular twin who did not have siblings would ipso facto fail to have nephews among his children. But I take it to be a priori obvious that the difference between having siblings and not having siblings does not constitute a difference in the causal powers of parents in virtue of its responsibility for this difference among the properties of their offspring.[9]

While we are on the topic of what is a priori obvious: the preceding three examples might suggest that the difference between $CP1$ and $CP2$ in instances of schema S does not constitute a difference of causal powers when $CP1$ and $CP2$ are *relational* properties. But a moment's reflection shows that this cannot be right (as, indeed, *Psychosemantics* was at some pains to point out). Taxonomy by relational properties is ubiquitous in the sciences, and it is not in dispute that properties like *being a meteor* or *being a planet*—properties which could, notice, distinguish molecularly identical chunks of rock—constitute causal powers. It is because *this* rock-twin is a planet and *that* rock-twin is not that this rock-twin has a Keplerian orbit and that rock-twin does not; it is because this rock-twin is a meteor and that rock-twin is not that this rock-twin's effects include craters and that rock-twin's effects do not. But, patently, *being a planet* and *being a meteor* are relational properties in good standing. To be a planet is to be a rock (or whatever) that is revolving around a star; to be a meteor is to be a rock (or whatever) that is falling, or has fallen, into collision with another rock.

---

[9] Another a priori taxonomic intuition that cries out to be taken seriously; it is (as *Psychosemantics* urged) preposterous to suggest that neurological (or biochemical; or molecular) states should be taxonomized by reference to the sorts of properties that distinguish twins in the standard examples—by whether there is water or twater in the local puddles, for example. Burge remarks that perhaps all this shows is that psychological taxonomy is more contextually sensitive than neurological taxonomy. But surely that will not run: if someone were to discover (for example) that living near high-tension wires turns your dendrites green, you can bet the neurologists would pay attention. So: what is the difference between living near high-tension wires and living near puddles of twater such that one, but not the other, is a candidate for a neurological causal power? (See below.)

Since the intuitions are pretty strong in all these cases, there is prima facie reason to suppose that there are some conditions on cause properties being causal powers which can be recognized from the armchair: not all cause properties are causal powers; not all relational properties fail to be, for two examples. I shall presently state a condition that, I claim, has to be satisfied if a property of a cause is a causal power in virtue of its responsibility for a certain property of an effect; and I shall claim that properties that distinguish twins (like being causally connected to water rather than twater; or having water thoughts rather than twater thoughts) do not satisfy this condition in virtue of their responsibility for differences between the intentional properties of the twins' behaviors. My evidence for the acceptability of this condition will be largely that it sorts examples like the ones I have just run through in an intuitively satisfactory way.

We need, however, some more ground clearing before we get down to it. (I solicit the reader's forbearance; once the ground is all clear, we shall be able to move very fast.) To begin with, I want to call your attention to what turns out to be a critical property of schema S: the question we are raising is not whether the difference between having $CP1$ and having $CP2$ is a difference in causal powers; rather, it is whether the difference between having $CP1$ and having $CP2$ is a difference in causal powers *in virtue of its being responsible for a certain difference between E1 and E2,* namely, in virtue of its being responsible for $E1$'s having $EP1$ rather than $EP2$ and for $E2$'s having $EP2$ rather than $EP1$. The point I am wanting to emphasize is that a cause property might fail to count as a causal power in virtue of its responsibility for one effect property, but still might constitute a causal power in virtue of its responsibility for some other effect property.

My coin's being heads up (rather than tails up) does not constitute a causal power in virtue of its responsibility for the universe being populated with H-particles; but it might constitute a causal power in virtue of its being responsible for my coin's reflecting light the way it does rather than some other way (e.g., rather than the way it would if it were tails up). Similarly, my having siblings does not constitute a causal power in virtue of its enabling me to have sons who are nephews. But suppose there is such a thing as sibling's disease; it causes people who have siblings to break out in a rash. Having a sibling might then be a causal power in virtue of being responsible for people coming down with sibling's disease. I stress this point because, if you do not relativize the issue about causal powers in this way, it turns out on a weak assumption that every contingent property is a causal power. The weak assumption is that it is (nomologically) possible to build a detector for any contingent property.

Consider, for example, the property of having had a Bulgarian grandmother who once plucked a daffodil as she went trippingly on her way to market. This is the sort of property that can distinguish between you and your molecular twin. And, no doubt, your effects have properties because of your having had such a grandmother which his do not have because he did not, e.g., your effects have one and all got the property of being the effects of someone who had a Bulgarian grandmother who . . ., etc. But, intuitively, that is not sufficient for making the possession of a Bulgarian *G* who . . .—or the lack of a Bulgarian *G* who . . .—a causal power. Indeed, you would not have thought that *any* of the Bulgarian-grandmother-dependent properties of your effects—any of the properties that distinguish the effects of the molecular twin with the Bulgarian *G* . . . from the effects of the molecular twin without one—would be of the right kind to make having a Bulgarian grandmother a causal power.

And yet, it would surely be possible to build (more precisely, it would surely have been possible to have built) a machine which exhaustively examines the piece of space-time that starts with the birth of your grandmother and ends with your birth and which goes into one state if it detects somebody who was your grandmother and was Bulgarian and once plucked a daffodil as she went trippingly on her way to market, but which goes into a different state in case it detects no such person. I assume that it is possible (in principle) to build a machine that reliably detects this property. I am prepared to assume that it is possible (in principle) to build such a machine for any contingent property at all. If this assumption is true, then the Bulgarian grandmother detector can distinguish between you and your molecular twin, and having a Bulgarian grandmother is having a causal power *in virtue of the (actual or possible) effects that instantiations of this property have on Bulgarian G . . . detectors.* (Quite generally, if any contingent property can be detected then any contingent property is a causal power in virtue of the effects of its instantiations on its detectors.) What does not follow, however, is that having had a Bulgarian *G* . . . is a causal power *in virtue of its effects on your behavior;* or, indeed, in virtue of its responsibility for *any property of yours.*

I am going through this song and dance by way of replying to an argument that Burge suggested (but does *not* endorse) that might lead one to think that having a water thought (rather than a twater thought) *must* be a causal power. After all, this argument says, it must be possible, in principle, to build a machine which looks through the chunk of space-time that starts with my birth and ends up with now, and which goes into one state if I am connected (in the right way) to water but goes into another state if I am connected (in the right way) to twater. This machine responds differently to me and my molecular twin, and does so in virtue of the fact that I am connected to water in the way that my twin is connected to *XYZ*. So my twin and I differ in our power to effect the states of this machine. So having water thoughts rather than twater thoughts (namely, having the kind of thoughts that you have if you are connected to water rather than twater) is a causal power in virtue of its responsibility for the machine's being in the state it is. So, if individualism says that having water thoughts rather than twater thoughts is *not* a causal power, then individualism is false.

Individualism does *not* say, however, that having water thoughts rather than twater thoughts is not a causal power. What it says is that having water thoughts rather than twater thoughts is not a causal power *in virtue of its being responsible for your producing water behaviors rather than twater behaviors.* (And similarly, being connected to water rather than twater is not a causal power in virtue of its being responsible for your having water thoughts rather than twater thoughts.) This is to say that the difference between having water thoughts and having twater thoughts is not a causal power in virtue of its responsibility for those of your properties *which are relevant to what psychological natural kinds your thoughts belong to,* e.g., it is not a causal power in virtue of its responsibility for the properties of your behavior.

That is all that an individualist has to show; to ask him to show more would be to make individualism false if there can be detectors for broad content properties. But that would be to trivialize the issue about individualism[10] since, as remarked above, it is plausible that there can be detectors for *any* contingent property (and, up to the limits of Turing machines, for quite a lot of noncontingent properties, too.)

So now: two causes differ in a certain property, and their effects differ in a certain property in virtue of this difference in the causes, and we want to know when the difference between the effects makes the difference between the causes a causal power. I shall tell you in just a moment. But it may be worth emphasizing that the cases we are interested in are not ones where the property that distinguishes the causes is *itself* the property of having a certain causal power.

Consider the case where *CP*1 just is the property of having the causal power to produce events that have *EP*1. Then, of course, the difference between having *CP*1 and not having it is the difference between having a certain causal power and not having it. This sort of case is easy because *it is noncontingent* that having *CP*1 is having a causal power. There are other, slightly less transparent, cases of this kind. For example, it is noncontingent that *being soluble in water* is having a causal power, because it is noncontingent that things that have that property have the power of dissolving in water; and the property of *being a camshaft* is a causal power because it is noncontingent that things that have that property have the power to lift the valves in a certain kind of engine (given conditions of optimal functioning . . ., etc.). There are, of course, scientifically interesting questions about properties that are noncontingently causal powers; for example, there are interesting questions about the supervenience bases of these properties and about their proper analysis into microfunctions. But there are not any interesting questions about whether things that have these properties have causal powers in virtue of having them. That question answers itself.

Compare, however, the properties of being a planet, being a meteor, and the like. These are causal powers in virtue of, for example, their respective abilities to produce Keplerian orbits and craters.

---

[10] Even being an H-particle would be a causal power in virtue of the nomological possibility of building H-particle detectors. Particle detecting is notoriously expensive and tricky; but once you have got a particle detector, it is trivial to convert it to detect H-particles. An H-particle detector is a particle detector that is hooked up to any gadget that is sensitive to which face of my coin is up. A particle detector that is hooked up to a Bulgarian grandmother would do.

But, insofar as being a planet is a causal power in virtue of the ability of planets to produce Keplerian orbits, it is contingent that being a planet is a causal power (for it is contingent that planets have Keplerian orbits); and insofar as being a meteor is a causal power in virtue of the ability of meteors to make craters, it is contingent that being a meteor is a causal power (for it is contingent that meteors make craters).

Notice that the broad content cases are like the meteor and planet cases and unlike the dispositional and functional cases. It may be that being connected to water rather than twater (hence having water thoughts rather than twater thoughts) is having a causal power; but if it is, it is contingent that it is. The property of being connected to water is not *identical* to the property of having a certain causal power, though it may be that there are causal powers that one has if one is connected to water that one would not have if one were not.

I draw two morals. First (contrary to some suggestions of R. Van Gulick's[11]), the fact that having a functional property is ipso facto having a certain causal power throws no particular light, one way or the other, on the question whether having a broad-content property is having a causal power. This conclusion may seem paradoxical since, after all, it is supposed to be that psychological properties *are* functional; so, if functional properties are noncontingently causal powers, and if psychological properties are functional, how could water thoughts and twater thoughts fail to be causal powers?

The answer is that it is actually not in dispute whether water thoughts and twater thoughts are causal powers. On the contrary, *of course* they are: my water thoughts are causally responsible for my reaching for water, my twin's twater thoughts are causally responsible for his reaching for twater . . . , and so on. The question on which local supervenience—hence individualism—turns, however, is whether the difference between water thoughts and twater thoughts is a *difference* in causal powers. The anti-individualist says 'yes, it is, in virtue of the intentional difference between the behaviors that water thoughts and twater thoughts cause'. The individualist says 'no it isn't. Water thoughts and twater thoughts are the same causal powers, only they're instantiated in people with different causal histories'. The idea that mental states are functional roles resolves this dispute only on the question-begging assumption that water behaviors and twater behaviors are behaviors of different kinds.

---

[11] "Metaphysical Arguments for Internalism and Why They Don't Work," in S. Silvers, ed., *Rerepresentation* (Philosophical Studies Series, #40) (Boston: Kluwer, 1989).

Imagine that we had different words for *being thirsty and born in the Bronx* and *being thirsty and born in Queens*. There would then be no question but that being Bronx-thirsty and being Queens-thirsty are causal powers; being Bronx-thirsty and being Queens-thirsty both make people drink, for example. But a question might arise whether being Bronx-thirsty and being Queens-thirsty are *different* causal powers.[12] My point is that this issue would still be open *even though* there is no question but that being Bronx-thirsty and being Queens-thirsty are causal powers.

A functionalist might undertake to argue that they are different causal powers because, on the one hand, psychological states are functionally individuated, and, on the other hand, Bronx thirst and Queens thirst lead to different behavioral consequences, namely, to Bronx thirst-quenching behaviors in the one case and Queens thirst-quenching behaviors in the other. But, clearly, to argue this way would be to beg the question. For, anybody who denies that Queens thirst and Bronx thirst are different causal powers, will also deny—and for the same reason—that Bronx thirst-quenching behavior and Queens thirst-quenching behavior are behaviors of different kinds. So, the principle that psychological states are functionally individuated does not settle questions about identity and difference of causal powers in this sort of case. And, of course, for an individualist, 'wants water' *is* this sort of case; for an individualist, 'wants water' means something like "thirsty and born *here*."

What distinguishes twins and threatens local supervenience is the property of *having water thoughts rather than twater thoughts*. No doubt, if functionalism is true, then mental states are one and all causal powers. But of course it does not follow from functionalism that differences between mental states are one and all functional differences; so it does not follow that differences between mental states are one and all differences of causal power. In fact, so far as I can tell, the only relevant connection between functionalism and the present issues about individualism is this: if, as I have claimed, the difference between the mental states of twins is at best *contingently* a difference of causal powers, it follows that it *cannot* be a difference of functional role, since differences of functional role are differences of causal powers *non*contingently.[13] Compare being water soluble rather than twater soluble. It is *not* contingent that having *this* property is having a causal power; to be soluble in water but not in twater just is to have the power to dissolve in the first but not in the second. So it is all right to allow that *being water soluble rather than twater soluble* is a functional property, as, indeed, intuition demands.

---

[12] This might well turn out to be the same question as whether 'is Bronx-thirsty' and 'is Queens-thirsty' are projectible; whether, for example, psychological generalizations about Bronx-thirsty people as such are confirmed by their instances.

[13] This comports with the idea that, whereas *having a belief* (desire, whatever) is being in the right functional state, having a belief *that P* is a matter of having the right head-to-world relations. So the difference between having a belief that P and a belief that Q *is not* a functional difference. For more on this, see *Psychosemantics*.

My impression is that this is all reasonably untendentious. In general, friends of broad content argue that it perfectly well *could turn out*—that there is no metaphysical reason why it should not turn out—that having mental states that differ in their broad content is having mental states that differ in their causal powers. (For example, it could turn out that there are causal laws that distinguish between twins.) But I do not remember hearing anyone argue that having mental states that differ in the way that the mental states of twins do *just is* having mental states that differ in their causal powers. On the contrary, according to the usual understanding, it is their causal *histories* that distinguish the mental states of twins. And the intuition about features of causal history is that some of them are causal powers (e.g., *having been dropped in transit; having been inoculated for smallpox*) and some of them are not (e.g., *having had a Bulgarian grandmother; having been born on a Tuesday*) and it is contingent which are which.

So, then, the second moral that I draw is that, in our hunt for a useful condition on what makes a difference in causal properties a difference in causal powers, we can restrict ourselves to cases where its *contingent* whether the difference in the properties constitutes a difference between powers. We are, finally, getting to where you can smell blood.

I am now, at last, going to tell you a story about why being a planet (for example) is a causal power and having siblings (for example) is not. I propose to do this in two steps. First, I shall tell you a simplified version of the story; it has the virtue of making the basic idea relatively transparent, but it has the disadvantage that it does not work. I shall then make the technical moves required to plug the leak; the complicated version that emerges will be a motivated condition upon a property of a cause being a causal power; one which, I claim, broad-content properties fail to meet.[14]

---

[14] Perhaps I had best reiterate that this is short for ". . . a motivated condition for a difference between properties of causes being a difference in their causal powers" and that the claim is not that being a water (/twater) thought is not a causal power, but rather that the difference between having a water thought and having a twater thought is not a difference in causal power More generally, the claim is that no two states differ in their causal powers *just* in virtue of differing in their broad contents.

So, then: here is me and here is my molecular twin; and I have siblings and he does not; and in virtue of my having siblings my sons are nephews, and in virtue of his not having siblings his sons are not nephews; and what we want to know is: Why is *having siblings* not a causal power in virtue of its being responsible for this difference in our offspring? Here is a first fling at the answer: it is because having siblings is *conceptually* connected to having sons who are nephews; to be a nephew *just is* to be a son whose parents have siblings. And, to put it roughly, your causal powers are a function of your *contingent* connections, not of your conceptual connections. As, indeed, Uncle Hume taught us.

Similarly: though it is a fact that all the world's particles become H-particles when my coin is heads up, that fact does not make being heads up a causal power of my coin. That is again because the connection between all the world's particles becoming H-particles at *t* and my coin's being heads up at *t* is conceptual. To be an H-particle at *t just is* to be a particle at a time when my coin is heads up.

Compare cases of relational properties that really are causal powers, like *being a planet*. Being a planet is a causal power in virtue of, for example, its contingent (a fortiori, nonconceptual) connection with having a Keplerian orbit. That is, being a planet is a causal power because it is true and contingent that, if you have molecularly identical chunks of rock, one of which is a planet and the other of which is not, then, ceteris paribus, the one which is a planet will have a Keplerian orbit, and ceteris paribus, the one which is not a planet will not.

Here is the general form of the proposed solution.[15] Consider an instance of schema *S*. *C*1 has *CP*1, *C*2 has *CP*2, *E*1 has *EP*1, *E*2 has *EP*2, and the difference between the causes is responsible for the difference between the effects in the sense that *E*1 would not have had *EP*1 (rather than *EP*2) but that *C*1 had *CP*1 (rather than *CP*2). And what we want to know is: When does the fact that this difference in the causes is responsible for this difference in the effects make *CP*1 and *CP*2 causal powers? The answer, which I shall call *condition C*, is:

> Only when it is not a conceptual truth that causes that which differ in that one has CP1 where the other has CP2 have effects that differ in that one has EP1 where the other has EP2.[16]

---

[15] But please, *please* keep it in mind that this solution applies in cases where, if the properties under consideration are causal powers, then they are causal powers contingently. In particular, it is *not* supposed to apply to properties that are causal powers necessarily, like dispositions. See above.

[16] I want to emphasize that condition *C* does not impugn the Davidsonian doctrine that the necessity/contingency of relations among *events* is description-rela-

So, for example, it is all right with condition *C* for ***being a meteor*** to be a causal power in virtue of the fact that meteors are responsible for craters. Take a pair of rock-twins such that one is a meteor and the other is not. Then (ceteris paribus) craters will be among the effects of the first but not among the effects of the second. And the relation between the difference between the rock-twins and the difference between their effects is nonconceptual. So it is all right for being a meteor (rather than a meteor twin) to be a causal power in virtue of the fact that meteors cause craters and meteor twins do not.[17]

Notice that the moral is that it is all right for being a meteor to be a causal power in virtue of this fact; not that being a meteor *is* a causal power in virtue of this fact. Satisfying condition *C* is, I suppose, necessary but not sufficient for being a causal power. As we are about to see, however, broad-content properties *fail* to satisfy this necessary condition, and that is enough to vindicate individualism.

Consider, first, the property of having water in your history (the property of being connected to water in whatever way it is that I am and my twin is not). The difference between being so connected and not being so connected is responsible for a certain difference between the broad contents of my thoughts and the broad contents of my twin's, namely, that I have water thoughts and he has twater thoughts. What we want to know is: Does this difference between our histories count as a causal power in virtue of the difference between the contents of our thoughts for which it is responsible. And the answer is: "No, because it is ***conceptually necessary*** that if you are connected to water in the right way then you have water thoughts (rather than twater thoughts) and it is again conceptually necessary that if you are connected to twater in the right way then you have

---

tive Suppose *e1* causes *e2*, so that '*e1* causes *e2*' is contingently true. Still, there will be descriptions satisfied by *e1* which entail (/presuppose) that *e1* causes *e2* (for example, such descriptions as 'the cause of *e2*'; 'the cause of *e2* caused *e2*' is, of course, necessarily true). But none of this implies that there is anything description-relative about whether the instantiation of one property entails the instantiation of another (about whether, for example, it is conceptually necessary that whatever instantiates bachelorhood instantiates unmarriedness; or whether it is conceptually necessary that whatever instantiates ***water-thinker-hood*** instantiates ***causally-connected-to-water-hood***. It is these latter sort of claims, not the former sort, which are at issue in respect of condition *C*.

[17] By contrast, I take it that their ability to produce ***meteor*** craters (where a meteor crater just is a crater that is caused by a meteor) is not a causal power that meteors have over and above their power to produce craters *tout court*. This is for the now familiar sort of reason: it is conceptually necessary that, whereas a crater caused (in the right way) by a meteor is a meteor crater, a crater caused (in whatever way) by a molecularly identical nonmeteor is not. Analogously, its ability to produce sunburns is not a causal power that the sun has over and above its power to produce burns.

twater thoughts (rather than water thoughts)." To have a water thought *just is* to have a thought that is connected to water in the right way, and to have a twater thought *just is* to have a thought that is connected to twater in the right way.[18] So it is not the case that my being connected to water rather than to twater is a difference in my causal powers in virtue of its responsibility for my having water thoughts rather than twater thoughts.

Now consider the difference between having water thoughts and having twater thoughts. Water thoughts cause water behavior (drilling for water and the like); twater thoughts cause twater behavior (drilling for twater and the like). What we want to know is: Does the difference between having water thoughts and having twater thoughts count as a causal power in virtue of the fact that it is responsible for this difference in the intentional properties of the behavior of the thinker? And the answer is: "No, because it is conceptually necessary that people who have water thoughts (rather than twater thoughts) produce water behavior (rather than twater behavior)." Being water behavior (rather than twater behavior) *just is* being behavior that is caused by water thoughts (rather than twater thoughts). So, though it is true that water thoughts are responsible for water behavior and twater thoughts are not, it does not follow that water thoughts have a causal power that twater thoughts lack. On the contrary, being a water thinker is the same causal power as being a twater thinker, only instantiated in a person with a different causal history.

I have been saying that it is only when the difference between the causes is not conceptually connected to the corresponding differences in the effects that the difference in the causes counts as a difference in causal powers. I think this really is the heart of the matter; it is why being responsible for particles being H-particles, being responsible for sons' being nephews, being responsible for behavior's being water behavior, and the like do not count as causal powers. But alas, the proposal does not work as stated and I shall have to do some patching.

Suppose that water is Bush's favorite thing to drink. Then, we have both:

1. If I am connected to water in the right way then my thoughts are water thoughts.

and

---

[18] More precisely (and assuming that functionalism is right about what beliefs, desires, and the like are) to have a water thought is to have a thought that (a) is connected to water in the right way; and (b) has whatever functional properties water thoughts and twater thoughts share. I leave out (b) in the text to simplify the exposition. No doubt, the exposition could do with some of that.

K-P 128

2. If I am connected to water in the right way, then my thoughts are thoughts about Bush's favorite drink.

Notice that both 1 and 2 distinguish me from my twin: because he is not connected to water in the right way, it is false of him that he has water thoughts, and it is also false of him that his thoughts are about *B*'s favorite drink. And, though 1 is conceptually necessary, 2 is contingent. That is, even though the difference between being water-connected and being twater-connected does not satisfy condition *C* in virtue of its being responsible for the difference between having water thoughts and having twater thoughts, it *does* satisfy condition *C* in virtue of its responsibility for the difference between having thoughts that are about Bush's favorite drink and having thoughts that are not. Hence, it satisfies condition *C* in virtue of making some prima facie *psychological* difference, some difference that is prima facie relevant to psychological taxonomy. So, even if, as I claim, condition *C* has its heart in the right place, still, as stated it has no teeth. Damn![19]

Everything is going to be all right, however. Let *G* be a property that nephews have, and let it be as contingent as you like that nephews have it. (*G* might be the property of glowing in the dark if it turns out that nephews do that.) Then the following is necessary:

3. If *G* is a property that nephews have, then if I have siblings then my sons have *G*.

So, for example, it is necessary that (if nephews glow in the dark, then if I have siblings, then my sons glow in the dark). The reason that this is necessary is, roughly, that it is conceptually necessary that the sons of people with siblings are nephews, so it is conceptually necessary that if you have a sibling then your sons have whatever properties nephews have. Notice, to repeat, that this is true even of properties that nephews have contingently: it is conceptually necessary that if *P* is a property that nephews have contingently then if you have siblings then your sons have *P*. (What *is not* true, of course, is 3′, the variant of 3 that has the modal operator imported. 3′ is false in the cases where *P* is a property that nephews have contingently.)

3′. If *P* is a property of nephews, then if you have siblings then it is necessary that your sons have *P*.

---

[19] I am indebted to Stich for this line of argument. I suppose he thinks I should be grateful.

Now compare 4, which I take to be clearly contingent.

4. If *E* is a property that Keplerian orbits have then, if I am a planet, then my orbit has *E*.

(You might wish to try out 4 reading *E* as the property of being eliptical.)

The reason that 4 is contingent is, roughly, that it is contingent that planets have Keplerian orbits, so it is contingent that if I am a planet then my orbit has whatever properties Keplerian orbits do. This is true even if *E* is a property that Keplerian orbits have necessarily (like being a Keplerian orbit).[20]

Now we can do the case that was bothering us before. Consider 5, where *B* might be the property of being a thought about Bush's favorite drink:

5. If *B* is a property that water thoughts have, then if I am connected to water in the right way, then *B* is a property that my thoughts have.

I take it that 5 is obviously conceptually necessary. The reason is, roughly, that it is conceptually necessary that if I am connected to water in the right way then my thoughts are water thoughts, and it is a truism that if something is a property of water thoughts then it is a property of my thoughts if my thoughts *are* water thoughts.[21] Similarly, mutatis mutandis, with 6.

6. If *B* is a property that water behaviors have, then if my thoughts are water thoughts then my behaviors have *B*.

6 is conceptually necessary because, roughly, it is conceptually necessary that the behaviors that water thoughts cause are water behaviors, so it is conceptually necessary that if something is a property of water behaviors then it is a property of the behaviors of water thinkers. This is true even if *B* is a property that water behaviors have contingently, like being behaviors that are concerned with Bush's favorite drink.

Compare these cases with ones where thoughts have bona fide causal powers. Suppose that thinking about topology gives one headaches. Then we have 7:

---

[20] 4 is necessary, however, if you choose *E* as some property that *everything* has necessarily; like being self-identical. This does not, of course, prejudice the present point.

[21] For those following the technicalities. in effect, my original line of argument depended on there being a conceptual relation between water connectedness and water thoughts construed *de dicto*. Stich gets a *contingent* relation between water connectedness and water thoughts *de re* by invoking such contingent premises as, for example, that water is Bush's favorite drink; and thus trivially satisfies condition *C*. In effect, 5 provides a substantive version of *C* by conditionalizing on these contingent premises.

7. If *B* is a property of headaches then if I have topology thoughts then my mental state has *B*.

7 is clearly contingent, and it is so because of the contingency of the (putative) relation between something's being a topology thought and its having headaches among its effects. Notice that 7 is contingent even if *B* is a property that headaches have necessarily (like being headaches).

So, here is the story. For the difference between being *CP*1 and being *CP*2 to be a difference of causal powers, it must at least be that the effects of being *CP*1 differ from the effects of being *CP*2. But, I claim, it is further required that this difference between the effects be *nonconceptually* related to the difference between the causes. This further condition is motivated both by our intuitions about the examples and by the Humean consideration that causal powers are, after all, powers to enter into nonconceptual relations. Broad-content differences, per se, do not satisfy this condition, however. There are differences between my behavior and my twin's which are due, in the first instance, to the difference between the intentional contents of our thoughts, and, in the second instance, to my being connected to water in a way that he is not. But these differences among the effects are conceptually related to the differences between the causes; it is conceptually necessary that being connected to water rather than twater leads to water thinking rather than twater thinking; and it is again conceptually necessary that water thinking leads to water behaving and twater thinking does not.

So, then, the difference between the mental states of twins does not count as a difference in causal power in virtue of its responsibility for the intentional differences among twin behaviors. So argument *B* is no good; what is wrong with it is that the inference from $2'$ to $3'$ is unsound.[22] Finally, since it is assumed that the effects of mental states that differ only in broad content are (relevantly) different *only*

---

[22] I can imagine that someone might now want to say. "All this follows from my conceding that the intentional properties that distinguish twins are causal powers contingently if they're causal powers at all. Well, I was wrong, those sorts of differences among broad content states are *non*contingently differences of causal powers. For example, the property of having water thoughts (rather than twater thoughts) is *identical* to the property of being able to produce water behaviors (rather than twater behaviors); the property of being connected to water (rather than twater) is *identical* to the property of being able to have water thoughts (rather than twater thoughts), etc. Notice that being soluble is a causal power even though the difference between being soluble and not is conceptually connected to the difference between dissolving and not."

All right, but it really does not help, if having water thoughts (rather than twater thoughts) is identical to having (inter alia) the power to drill for water (rather than twater), then the broad-content psychological generalizations that distinguish twins under intentional description, it follows that there are *no* taxonomically relevant differences consequent upon broad-content differences as such. From the point of view of psychological taxonomy, my mental states must therefore belong to the same natural kind as those of my molecular twin. So individualism is true and local supervenience is preserved. End of story.

—Oi! Halt thief! Stop that person.
—What's the matter?
—You promised me an argument *for* narrow content. But all you've given me is an argument *against* arguments against individualism. I want my money back.
—Sorry.

We have seen that twater thoughts and water thoughts are not different causal powers. So, for the psychologist's purposes, they are the same intentional state.[23] But they cannot be the same intentional state unless they have the same intentional content. And they cannot have the same intentional content unless intentional content is individuated narrowly. Now it is an argument for narrow content.

—Much obliged.
—My pleasure.

JERRY A FODOR

Rutgers University and CUNY Graduate Center

APPENDIX: A NOTE ON THE CROSS-CONTEXT TEST FOR IDENTITY OF CAUSAL POWERS

Suppose that my twin and I are both green, but I live in red world (where everything except me is red) and he lives in green world (where everything including him is green). So then, he has a property of ***being adapted to the***

[like, 'if you have water wants (rather than twater wants) then you drill for water (rather than twater)'] themselves all come out conceptually necessary (a kind of point to which Ryleans were, of course, entirely alert). So, if the assumption is that intentional states are noncontingently causal powers, then the appropriate form for the individualist's supervenience claim is that no *contingent* intentional generalization can distinguish twins (no contingent intentional generalization can be such that one but not the other of the twins satisfies its antecedent or consequent). The moral would then be: no causal laws about broad intentional states as such (in fact, just the sort of moral that Ryleans used to draw).

It is true that being soluble is a causal power even though it is conceptually connected to dissolving. But the price for thus evading condition *C* is the "quasi-logical" status of 'if soluble then dissolves'.

[23] —But why can this putative psychologist not allow them to be *different* states but with the *same* causal powers?

—Because, if he does, his theory misses generalizations, namely, all the generalizations that subsume me and my twin. Good taxonomy is about *not* missing generalizations.

color of his environment which I lack. And, intuitively, this property is a causal power: because he has it, he can sneak up on things; because I lack it, I cannot. But this appears to make trouble for the idea that identity of causal powers has to be assessed *across* contexts. The cross-context test implies that there cannot be a difference in our causal powers unless there are counterfactuals that are true of one of us but not of the other; but, in the present case, it looks like the counterfactuals come out the same for him as they do for me. If he were in my world, he could not sneak up on things; if I were in his world, I could. So our causal powers are the same according to the cross-context test. So there must be something wrong with the idea that identity of causal powers should be assessed across contexts. (For elaboration of what I take to be essentially this line of argument, see Van Gulick, *op. cit.*)

But this argument is fallacious; it rests on a confusion between, to put it roughly, an attributive and a referential way of reading the definite description in 'the color of his environment', with a consequent confusion about how the across-context test applies. This will only take a minute to sort out. If you read 'the color of his environment' referentially, then my twin's being adapted to the color of his environment is his being green. Read that way, then, *I too am adapted to the color of his environment* (I am green, too); so being adapted to the color of his environment is a causal power that we *share*. As, indeed, the cross-context test for causal powers properly predicts: if I were in his environment, I would be able to sneak up on things; if he were in mine, he would not. It is just that, since I am *not* in his environment and he is, being adapted to the color of his environment does him a lot more good than it does me.

If, however, you read 'the color of his environment' attributively, then his being adapted to the color of his environment consists in his being the *same color as the environment that he is in*. So, if you move *that* property across contexts, then when he is in *my* environment his being adapted to the color of his environment consists in his being (not green like me but) *red*. So, if he is adapted to the color of his environment in *that* sense, then when we are both in my environment, he can sneak up on things and I cannot. So, according to the across-context test, being adapted to the color of his environment is a causal power that my twin has and I lack, which is, again, the intuitively correct result.

So there is nothing wrong with the test.

(—Well, but what about the property of *being-adapted-to-the-color-of-red-world-and-in-red-world*. Is *that* not a causal power that defies the cross-context test?

—Answer: It does not *defy* the cross-context test; it merely satisfies it trivially. Red-world is the only context in which one *can* have that property.)

REAL PATTERNS*

ARE there really beliefs? Or are we learning (from neuroscience and psychology, presumably) that, strictly speaking, beliefs are figments of our imagination, items in a superseded ontology? Philosophers generally regard such ontological questions as admitting just two possible answers: either beliefs exist or they do not. There is no such state as quasi existence; there are no stable doctrines of semirealism. Beliefs must either be vindicated along with the viruses or banished along with the banshees. A bracing conviction prevails, then, to the effect that when it comes to beliefs (and other mental items) one must be either a realist or an eliminative materialist.

I REALISM ABOUT BELIEFS

This conviction prevails in spite of my best efforts over the years to undermine it with various analogies: are *voices* in your ontology?[1] Are *centers of gravity* in your ontology?[2]

It is amusing to note that my analogizing beliefs to centers of gravity has been attacked from both sides of the ontological dichotomy, by philosophers who think it is simply obvious that centers of gravity are useful fictions, and by philosophers who think it is simply obvious that centers of gravity are perfectly real:

> The trouble with these supposed parallels . . . is that they are all strictly speaking *false*, although they are no doubt useful simplifications for many purposes. It is false, for example, that the gravitational attraction between the Earth and the Moon involves two point masses; but it is a good enough first approximation for many calculations. However, this is not at all what Dennett really wants to say about intentional states. For he insists that to adopt the intentional stance and interpret an agent as acting on certain beliefs and desires is to discern a pattern in his actions which is genuinely there (a pattern which is missed if we instead adopt a scientific stance): Dennett certainly does not hold that the role of intentional ascriptions is merely to give us a useful approximation to a truth that can be more accurately expressed in non-intentional terms.[3]

---

* Thanks to Kathleen Akins, Akeel Bilgrami, Donald Davidson, Barbara Hannan, Douglas Hofstadter, Norton Nelkin, W. V. O. Quine, Richard Rorty, George Smith, Peter Suber, Stephen White, and the MIT/Tufts philosophy of psychology discussion group for the discussions that provoked and shaped this paper.

[1] *Content and Consciousness* (Boston: Routledge & Kegan Paul, 1969), ch. 1.

[2] "Three Kinds of Intentional Psychology," in R. Healey, ed., *Reduction, Time and Reality* (New York: Cambridge, 1981); and *The Intentional Stance* (Cambridge: MIT, 1987).

[3] Peter Smith, "Wit and Chutzpah," review of *The Intentional Stance* and Jerry A. Fodor's *Psychosemantics*, *Times Higher Education Supplement* (August 7, 1988), p. 22.

0022-362X/91/8801/27–51

Compare this with Fred Dretske's[4] equally confident assertion of realism:

> I am a realist about centers of gravity. . . . The earth obviously exerts a gravitational attraction on *all* parts of the moon—not just its center of gravity. The *resultant* force, a vector sum, acts through a point, but this is something quite different. One should be very clear about what centers of gravity are *before* deciding whether to be literal about them, *before* deciding whether or not to be a center-of-gravity realist (*ibid.*, p. 511).

Dretske's advice is well-taken. What are centers of gravity? They are mathematical points—abstract objects or what Hans Reichenbach called *abstracta*—definable in terms of physical forces and other properties. The question of whether abstract objects are real —the question of whether or not "one should be a realist about them"—can take two different paths, which we might call the metaphysical and the scientific. The metaphysical path simply concerns the reality or existence of abstract objects generally, and does not distinguish them in terms of their scientific utility. Consider, for instance, the *center of population* of the United States. I define this as the mathematical point at the intersection of the two lines such that there are as many inhabitants north as south of the latitude, and as many inhabitants east as west of the longitude. This point is (or can be) just as precisely defined as the center of gravity or center of mass of an object. (Since these median strips might turn out to be wide, take the midline of each strip as the line; count as inhabitants all those within the territorial waters and up to twenty miles in altitude —orbiting astronauts do not count—and take each inhabitant's navel to be the determining point, etc.) I do not know the center of population's current geographic location, but I am quite sure it is west of where it was ten years ago. It jiggles around constantly, as people move about, taking rides on planes, trains, and automobiles, etc. I doubt that this abstract object is of any value at all in any scientific theory, but just in case it is, here is an even more trivial abstract object: Dennett's lost sock center: the point defined as the center of the smallest sphere that can be inscribed around all the socks I have ever lost in my life.

These abstract objects have the same metaphysical status as centers of gravity. Is Dretske a realist about them all? Should we be? I do not intend to pursue this question, for I suspect that Dretske is—and we should be—more interested in the scientific path to realism: centers of gravity are real because they are (somehow) *good* abstract objects. They deserve to be taken seriously, learned about, used. If we go so far as to distinguish them as *real* (contrasting them, perhaps, with those abstract objects which are *bogus*), that is because we think they serve in perspicuous representations of real forces, "natural" properties, and the like. This path brings us closer, in any case, to the issues running in the debates about the reality of beliefs.

---

[4] "The Stance Stance," commentary on *The Intentional Stance*, in *Behavioral and Brain Sciences*, XI (1988): 511–2.

I have claimed that beliefs are best considered to be abstract objects rather like centers of gravity. Smith considers centers of gravity to be useful fictions while Dretske considers them to be useful (and hence?) real abstractions, and each takes his view to constitute a criticism of my position. The optimistic assessment of these opposite criticisms is that they cancel each other out; my analogy must have hit the nail on the head. The pessimistic assessment is that more needs to be said to convince philosophers that a mild and intermediate sort of realism is a positively attractive position, and not just the desperate dodge of ontological responsibility it has sometimes been taken to be. I have just such a case to present, a generalization and extension of my earlier attempts, via the concept of a *pattern*. My aim on this occasion is not so much to prove that my intermediate doctrine about the reality of psychological states is right, but just that it is quite possibly right, because a parallel doctrine is demonstrably right about some simpler cases.

We use folk psychology—interpretation of each other as believers, wanters, intenders, and the like—to predict what people will do next. Prediction is not the only thing we care about, of course. Folk psychology helps us understand and empathize with others, organize our memories, interpret our emotions, and flavor our vision in a thousand ways, but at the heart of all these is the enormous predictive leverage of folk psychology. Without its predictive power, we could have no interpersonal projects or relations at all; human activity would be just so much Brownian motion; we would be baffling ciphers to each other and to ourselves—we could not even conceptualize our own flailings. In what follows, I shall concentrate always on folk-psychological prediction, not because I make the mistake of ignoring all the other interests we have in people aside from making bets on what they will do next, but because I claim that our power to *interpret* the actions of others depends on our power—seldom explicitly exercised—to predict them.[5]

---

[5] R. A. Sharpe, in "Dennett's Journey Towards Panpsychism," *Inquiry*, XXXII (1989): 233–240, takes me to task on this point, using examples from Proust to drive home the point that "Proust draws our attention to possible lives and these

Where utter patternlessness or randomness prevails, nothing is predictable. The success of folk-psychological prediction, like the success of any prediction, depends on there being some order or pattern in the world to exploit. Exactly where in the world does this pattern exist? What is the pattern a pattern *of*?[6] Some have thought, with Fodor, that the pattern of belief must in the end be a pattern of structures in the brain, formulae written in the language of thought. Where else could it be? Gibsonians might say the pattern is "in the light"—and Quinians (such as Donald Davidson and I) could almost agree: the pattern is discernible in agents' (observable) behavior when we subject it to "radical interpretation" (Davidson) "from the intentional stance" (Dennett).

When are the elements of a pattern real and not merely apparent? Answering this question will help us resolve the misconceptions that have led to the proliferation of "ontological positions" about beliefs, the different grades or kinds of realism. I shall concentrate on five salient exemplars arrayed in the space of possibilities: Fodor's industrial-strength Realism (he writes it with a capital 'R'); Davidson's regular strength realism; my mild realism; Richard Rorty's milder-than-mild irrealism, according to which the pattern is *only* in the eyes of the beholders, and Paul Churchland's eliminative materialism, which denies the reality of beliefs altogether.

In what follows, I shall assume that these disagreements all take place within an arena of common acceptance of what Arthur Fine[7] calls NOA, the natural ontological attitude. That is, I take the interest in these disagreements to lie not in differences of opinion about the ultimate metaphysical status of physical things or abstract things (e.g., electrons or centers of gravity), but in differences of opinion about whether beliefs and other mental states are, shall we say, *as real as* electrons or centers of gravity. I want to show that mild realism is the doctrine that makes the most sense when what we are

---

possible lives are various. But in none of them is prediction of paramount importance" (240). I agree. I also agree that what makes people interesting (in novels and in real life) is precisely their unpredictability But that unpredictability is only interesting against the backdrop of routine predictability on which all interpretation depends As I note in *The Intentional Stance* (p. 79) in response to a similar objection of Fodor's, the same is true of chess: the game is interesting only because of the unpredictability of one's opponent, but that is to say: the intentional stance can usually eliminate *only* ninety percent of the legal moves.

[6] Norton Nelkin, "Patterns," forthcoming.

[7] *The Shaky Game: Einstein Realism and the Quantum Theory* (Chicago University Press, 1986); see esp. p. 153n, and his comments there on Rorty, which I take to be consonant with mine here.

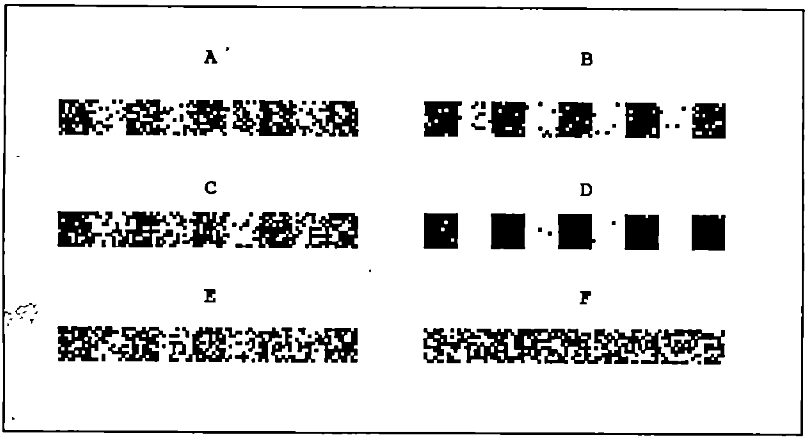


Figure 1

talking about is real patterns, such as the real patterns discernible from the intentional stance.[8]

In order to make clear the attractions and difficulties of these different positions about patterns, I shall apply them first to a much simpler, more readily visualized, and uncontroversial sort of pattern.

II THE REALITY OF PATTERNS

Consider the six objects in Figure 1 (which I shall call *frames*):

We can understand a frame to be a finite subset of data, a window on an indefinitely larger world of further data. In one sense *A–F* all display different patterns; if you look closely you will see that no two frames are exactly alike ("atom-for-atom replicas," if you like). In another sense, *A–F* all display the same pattern; they were all made by the same basic process, a printing of ten rows of ninety dots, ten black dots followed by ten white dots, etc. The overall effect is to create five equally spaced black squares or bars in the window. I take it that this pattern, which I shall dub *bar code,* is a real pattern if anything is. But some random (actually pseudo-random) "noise" has been allowed to interfere with the actual printing. The noise ratio is as follows:

A: 25%   *B*: 10%
C: 25%   *D*: 1%
E: 33%   *F*: 50%

---

[8] See *The Intentional Stance,* pp. 38–42, "Real patterns, deeper facts, and empty questions."

It is impossible to see that *F* is not purely (pseudo-) random noise; you will just have to take my word for it that it was actually generated by the same program that generated the other five patterns; all I changed was the noise ratio.

Now, what does it mean to say that a pattern in one of these frames is real, or that it is really there? Given our privileged information about how these frames were generated, we may be tempted to say that there is a single pattern in all six cases—even in *F*, where it is "indiscernible." But I propose that the self-contradictory air of "indiscernible pattern" should be taken seriously. We may be able to make some extended, or metaphorical, sense of the idea of indiscernible patterns (or invisible pictures or silent symphonies), but in the root case a pattern is "by definition" a candidate for pattern *recognition*. (It is this loose but unbreakable link to observers or perspectives, of course, that makes "pattern" an attractive term to someone perched between instrumentalism and industrial-strength realism.)

Fortunately, there is a standard way of making these intuitions about the discernibility-in-principle of patterns precise. Consider the task of transmitting information about one of the frames from one place to another. How many bits of information will it take to transmit each frame? The least efficient method is simply to send the "bit map," which identifies each dot *seriatim* ("dot one is black, dot two is white, dot three is white, . . ."). For a black-and-white frame of 900 dots (or pixels, as they are called), the transmission requires 900 bits. Sending the bit map is in effect verbatim quotation, accurate but inefficient. Its most important virtue is that it is equally capable of transmitting any pattern or any particular instance of utter patternlessness.

Gregory Chaitin's[9] valuable definition of mathematical randomness invokes this idea. A series (of dots or numbers or whatever) is random if and only if the information required to describe (transmit) the series accurately is *incompressible:* nothing shorter than the verbatim bit map will preserve the series. Then a series is not random—has a pattern—if and only if there is some more efficient way of describing it.[10] Frame *D*, for instance, can be described as "ten rows of ninety: ten black followed by ten white, etc., *with the following exceptions:* dots 57, 88, ..." This expression, suitably encoded, is much shorter than 900 bits long. The comparable expressions for the other frames will be proportionally longer, since they will have to mention, verbatim, more exceptions, and the degeneracy of the "pattern" in F is revealed by the fact that its description in this system will be no improvement over the bit map—in fact, it will tend on average to be trivially longer, since it takes some bits to describe the pattern that is then obliterated by all the exceptions.

---

[9] "Randomness and Mathematical Proof," *Scientific American*, CCXXXII (1975): 47–52.

[10] More precisely. "A series of numbers is random if the smallest algorithm capable of specifying it to a computer has about the same number of bits of information as the series itself" (Chaitin, p. 48). This is what explains the fact that the "random number generator" built into most computers is not really properly named, since it is some function describable in a few bits (a little subroutine that is called for some output whenever a program requires a "random" number or series). If I send you the description of the pseudo-random number generator on

Of course, there are bound to be other ways of describing the evident patterns in these frames, and some will be more efficient than others—in the precise sense of being systematically specifiable in fewer bits.[11] Any such description, if an improvement over the bit map, is the description of a real pattern in the data.[12]

Consider bar code, the particular pattern seen in A–E, and almost perfectly instantiated in D. *That* pattern is quite readily discernible to the naked human eye in these presentations of the data, because of the particular pattern-recognition machinery hard-wired in our visual systems—edge detectors, luminance detectors, and the like. But the very same data (the very same streams of bits) presented in some other format might well yield no hint of pattern to us, especially in the cases where bar code is contaminated by salt and pepper, as in frames A through C. For instance, if we broke the 900-bit series of frame B into 4-bit chunks, and then translated each of these into hexadecimal notation, one would be hard pressed indeed to tell the resulting series of hexadecimal digits from a random series, since the hexadecimal chunking would be seriously out of phase with the deci-

---

my computer, you can use it to generate exactly the same infinite series of random-seeming digits.

[11] Such schemes for efficient description, called compression algorithms, are widely used in computer graphics for saving storage space. They break the screen into uniformly colored regions, for instance, and specify region boundaries (rather like the "paint by numbers" line drawings sold in craft shops). The more complicated the picture on the screen, the longer the compressed description will be; in the worst case (a picture of confetti randomly sprinkled over the screen) the compression algorithm will be stumped, and can do no better than a verbatim bit map.

[12] What about the "system" of pattern description that simply baptizes frames with proper names (A through F, in this case) and tells the receiver which frame is up by simply sending 'F'? This looks much shorter than the bit map until we consider that such a description must be part of an entirely general system. How many proper names will we need to name all possible 900-dot frames? Trivially, the 900-bit binary number, 11111111 . . . . To send the "worst-case" proper name will take exactly as many bits as sending the bit map. This confirms our intuition that proper names are maximally inefficient ways of couching generalizations ("Alf is tall and Bill is tall and . . .").

mal pattern—and hence the "noise" would not "stand out" as noise. There are myriad ways of displaying any 900-bit series of data points, and not many of them would inspire us to concoct an efficient description of the series. Other creatures with different sense organs, or different interests, might readily perceive patterns that were imperceptible to us. The patterns would be *there* all along, but just invisible to *us*.

The idiosyncracy of perceivers' capacities to discern patterns is striking. Visual patterns with axes of vertical symmetry stick out like sore thumbs for us, but if one simply rotates the frame a few degrees, the symmetry is often utterly beyond noticing. And the "perspectives" from which patterns are "perceptible" are not restricted to variations on presentation to the sense modalities. Differences in knowledge yield striking differences in the capacity to pick up patterns. Expert chess players can instantly perceive (and subsequently recall with high accuracy) the total board position in a real game, but are much worse at recall if the same chess pieces are randomly placed on the board, even though to a novice both boards are equally hard to recall.[13] This should not surprise anyone who considers that an expert speaker of English would have much less difficulty perceiving and recalling

> The frightened cat struggled to get loose.

than

> Te ser.ioghehnde t srugfcalde go tgtt ohle

which contains the same pieces, now somewhat disordered. Expert chess players, unlike novices, not only know how to *play* chess; they know how to *read* chess—how to see the patterns at a glance.

A pattern exists in some data—is real—if *there is* a description of the data that is more efficient than the bit map, whether or not anyone can concoct it. Compression algorithms, as general-purpose pattern describers, are efficient ways of transmitting exact copies of frames, such as *A–F*, from one place to another, but our interests often favor a somewhat different goal: transmitting *inexact* copies that nevertheless preserve "the" pattern that is important to us. For some purposes, we need not list the exceptions to bar code, but only transmit the information that the pattern is bar code with *n*% noise. Following this strategy, frames *A* and *C*, though discernibly different under careful inspection, count as *the same pattern*, since what matters to us is that the pattern is bar code with 25% noise, and we do not care which particular noise occurs, only that it occurs.

---

[13] A. D. de Groot, *Thought and Choice in Chess* (The Hague: Mouton, 1965)

Sometimes we are interested in not just ignoring the noise, but eliminating it, improving the pattern in transmission. Copy-editing is a good example. Consider the likely effoct thes santince wull hive hod on tha cupy adutor whu preparis thas monescrupt fur prunteng. *My* interest in this particular instance is that the "noise" be transmitted, not removed, though I actually do not care exactly *which* noise is there.

Here then are three different attitudes we take at various times toward patterns. Sometimes we care about exact description or reproduction of detail, at whatever cost. From this perspective, a real pattern in frame *A* is *bar code with the following exceptions: 7, 8, 11, ...*. At other times we care about the noise, but not where in particular it occurs. From this perspective, a real pattern in frame *A* is *bar code with 25% noise.* And sometimes, we simply tolerate or ignore the noise. From this perspective, a real pattern in frame *A* is simply: *bar code.* But is bar code really there in frame A? I am tempted to respond: Look! You can see it with your own eyes. But there is something more constructive to say as well.

When two individuals confront the same data, they may perceive different patterns in them, but since we can have varied interests and perspectives, these differences do not all count as disagreements. Or in any event they should not. If Jones sees pattern α (with *n*% noise) and Brown sees pattern β (with *m*% noise) there may be no ground for determining that one of them is right and the other wrong. Suppose they are both using their patterns to bet on the next datum in the series. Jones bets according to the "pure" pattern α, but budgets for *n*% errors when he looks for odds. Brown does likewise, using pattern β. If both patterns are real, they will both get rich. That is to say, so long as they use their expectation of deviations from the "ideal" to temper their odds policy, they will do better than chance —perhaps very much better.

Now suppose they compare notes. Suppose that α is a simple, easy-to-calculate pattern, but with a high noise rate—for instance, suppose α is bar code as it appears in frame *E*. And suppose that Brown has found some periodicity or progression in the "random" noise that Jones just tolerates, so that β is a much more complicated description of pattern-superimposed-on-pattern. This permits Brown to do better than chance, we may suppose, at predicting when the "noise" will come. As a result, Brown budgets for a lower error rate—say only 5%. "What you call noise, Jones, is actually pattern," Brown might say. "Of course there is still *some* noise in my pattern, but my pattern is better—more real—than yours! Yours is actually just a mere appearance." Jones might well reply that it is all a matter of taste; he notes how hard Brown has to work to calculate predictions, and points to the fact that he is getting just as rich (or maybe richer) by using a simpler, sloppier system and making more bets at good odds than Brown can muster. "My pattern is perfectly real—look how rich I'm getting. If it were an illusion, I'd be broke."

This crass way of putting things—in terms of betting and getting rich—is simply a vivid way of drawing attention to a real, and far from crass, trade-off that is ubiquitous in nature, and hence in folk psychology. Would we prefer an extremely compact pattern description with a high noise ratio or a less compact pattern description with a lower noise ratio? Our decision may depend on how swiftly and reliably we can discern the simple pattern, how dangerous errors are, how much of our resources we can afford to allocate to detection and calculation. These "design decisions" are typically not left to us to make by individual and deliberate choices; they are incorporated into the design of our sense organs by genetic evolution, and into our culture by cultural evolution. The product of this design evolution process is what Wilfrid Sellars[14] calls our *manifest image,* and it is composed of folk physics, folk psychology, and the other pattern-making perspectives we have on the buzzing blooming confusion that bombards us with data. The ontology generated by the manifest image has thus a deeply pragmatic source.[15]

Do these same pragmatic considerations apply to the scientific image, widely regarded as the final arbiter of ontology? Science is supposed to carve nature at the joints—at its *real* joints, of course. Is it permissible in science to adopt a carving system so simple that it makes sense to tolerate occasional misdivisions and consequent mispredictions? It happens all the time. The ubiquitous practice of using idealized models is exactly a matter of trading off reliability and accuracy of prediction against computational tractability. A particularly elegant and handy oversimplification may under some circumstances be irresistible. The use of Newtonian rather than Einsteinian mechanics in most mundane scientific and engineering calculations is an obvious example. A tractable oversimplification may be attractive even in the face of a high error rate; considering inherited traits to be carried by single genes "for" those traits is an example; considering agents in the marketplace to be perfectly rational self-aggrandizers with perfect information is another.

---

[14] *Science, Perception and Reality* (Boston: Routledge & Kegan Paul, 1963).

[15] In "Randomness and Perceived Randomness in Evolutionary Biology," *Synthese,* XLIII (1980): 287–329, William Wimsatt offers a nice example (296): while the insectivorous bird tracks individual insects, the anteater just averages over the ant-infested area; one might say that, while the bird's manifest image quantifies over insects, 'ant' is a mass term for anteaters. See the discussion of this and related examples in my *Elbow Room* (Cambridge: MIT, 1984), pp. 108–110

III. PATTERNS IN LIFE

The time has come to export these observations about patterns and reality to the controversial arena of belief attribution. The largish leap we must make is nicely expedited by pausing at a stepping-stone example midway between the world of the dot frames and the world of folk psychology: John Horton Conway's Game of Life. In my opinion, every philosophy student should be held responsible for an intimate acquaintance with the Game of Life. It should be considered an essential tool in every thought-experimenter's kit, a prodigiously versatile generator of philosophically important examples and thought experiments of admirable clarity and vividness. In *The Intentional Stance,* I briefly exploited it to make a point about the costs and benefits of risky prediction from the intentional stance,[16] but I have since learned that I presumed too much familiarity with the underlying ideas. Here, then, is a somewhat expanded basic introduction to Life.[17]

Life is played on a two-dimensional grid, such as a checkerboard or a computer screen; it is not a game one plays to win; if it is a game at all, it is solitaire. The grid divides space into square cells, and each cell is either ON or OFF at each moment. Each cell has eight neighbors: the four adjacent cells north, south, east, and west, and the four diagonals: northeast, southeast, southwest, and northwest. Time in the Life world is also discrete, not continuous; it advances in ticks, and the state of the world changes between each tick according to the following rule:

> Each cell, in order to determine what to do in the next instant, counts how many of its eight neighbors is ON at the present instant. If the answer is exactly two, the cell stays in its present state (ON or OFF) in the next instant. If the answer is exactly three, the cell is ON in the next instant whatever its current state. Under all other conditions the cell is OFF.

The entire physics of the Life world is captured in that single, unexceptioned law. [While this is the fundamental law of the "physics" of the Life world, it helps at first to conceive this curious physics in biological terms: think of cells going ON as births, cells going OFF as deaths, and succeeding instants as generations. Either overcrowding (more than three inhabited neighbors) or isolation (less than two inhabited neighbors) leads to death.] By the scrupulous application of this single law, one can predict with perfect accuracy the next instant of any configuration of ON and OFF cells, and the instant after that, and so forth. In other words, the Life world is a toy world that perfectly instantiates Laplace's vision of determinism: given the state description of this world at an instant, we finite observers can perfectly predict the future instants by the simple application of our one law of physics. Or, in my terms, when we adopt the physical stance toward a configuration in the Life world, our powers of prediction are perfect: there is no noise, no uncertainty, no probability less than one. Moreover, it follows from the two-dimensionality of the Life world that nothing is hidden from view. There is no backstage; there are no hidden variables; the unfolding of the physics of objects in the Life world is directly and completely visible.

---

[16] *The Intentional Stance,* pp. 37–9.

[17] Martin Gardner introduced the Game of Life to a wide audience in two columns in *Scientific American* in October, 1970, and February, 1971. William Poundstone, *The Recursive Universe: Cosmic Complexity and the Limits of Scientific Knowledge* (New York: Morrow, 1985), is an excellent exploration of the game and its philosophical implications. Two figures from Poundstone's book are reproduced, with kind permission from the author and publisher, on pp. 39 & 40.

There are computer simulations of the Life world in which one can set up configurations on the screen and then watch them evolve according to the single rule. In the best simulations, one can change the scale of both time and space, alternating between close-up and bird's-eye view. A nice touch added to some color versions is that ON cells (often just called pixels) are color-coded by their age; they are born blue, let us say, and then change color each generation, moving through green to yellow to orange to red to brown to black and then staying black unless they die. This permits one to see at a glance how old certain patterns are, which cells are co-generational, where the birth action is, and so forth.[18]

One soon discovers that some simple configurations are more interesting than others. In addition to those configurations which never change—the "still lifes" such as four pixels in a square—and those which evaporate entirely—such as any long diagonal line segment, whose two tail pixels die of isolation each instant until the line disappears entirely—there are configurations with all manner of periodicity. Three pixels in a line make a simple flasher, which becomes three pixels in a column in the next instant, and reverts to three in a line in the next, ad infinitum, unless some other configuration encroaches. Encroachment is what makes Life interesting:

---

[18] Poundstone, *op. cit.*, provides simple BASIC and IBM-PC assembly language simulations you can copy for your own home computer, and describes some of the interesting variations.

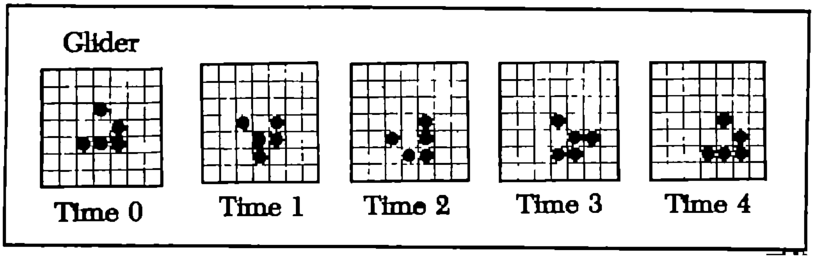


Figure 2: (from Poundstone, *op. cit.*).

among the periodic configurations are some that swim, amoeba-like, across the plane. The simplest is the *glider*, the five-pixel configuration shown taking a single stroke to the southeast in figure 2. Then there are the eaters, the puffer trains, and space rakes, and a host of other aptly named denizens of the Life world that emerge in the ontology of a new level, analogous to what I have called the design level. This level has its own language, a transparent foreshortening of the tedious descriptions one could give at the physical level. For instance:

> An eater can eat a glider in four generations. Whatever is being consumed, the basic process is the same. A bridge forms between the eater and its prey. In the next generation, the bridge region dies from overpopulation, taking a bite out of both eater and prey. The eater then repairs itself. The prey usually cannot. If the remainder of the prey dies out as with the glider, the prey is consumed (*ibid.*, p. 38).

Note that there has been a distinct ontological shift as we move between levels; whereas at the physical level there is no motion, and the only individuals, cells, are defined by their fixed spatial location, at this design level we have the motion of persisting objects; it is one and the same glider that has moved southeast in figure 2, changing shape as it moves, and there is one less glider in the world after the eater has eaten it in figure 3. (Here is a warming-up exercise for what is to follow: should we say that there is *real* motion in the Life world, or only *apparent* motion? The flashing pixels on the computer screen are a paradigm case, after all, of what a psychologist would call apparent motion. Are there *really* gliders that move, or are there just patterns of cell state that move? And if we opt for the latter, should we say at least that these moving patterns are real?)

Notice, too, that at this level one proposes generalizations that require 'usually' or 'provided nothing encroaches' clauses. Stray bits of debris from earlier events can "break" or "kill" one of the objects

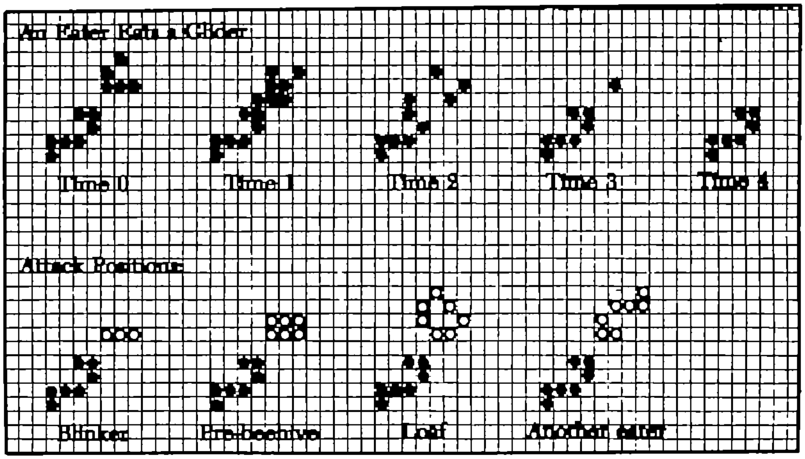


Figure 3 (from Poundstone, *op. cit.*)

in the ontology at this level; their *salience as real things* is considerable, but not guaranteed. To say that their salience is considerable is to say that one can, with some small risk, ascend to this design level, adopt its ontology, and proceed to predict—sketchily and riskily—the behavior of larger configurations or systems of configurations, without bothering to compute the physical level. For instance, one can set oneself the task of designing some interesting supersystem out of the "parts" that the design level makes available. Surely the most impressive triumph of this design activity in the Life world is the proof that a working model of a universal Turing machine can in principle be constructed in the Life plane! Von Neumann had already shown that in principle a two-dimensional universal Turing machine could be constructed out of cellular automata, so it was "just" a matter of "engineering" to show how, in principle, it could be constructed out of the simpler cellular automata defined in the Life world. Glider streams can provide the tape, for instance, and the tape reader can be some huge assembly of eaters, gliders, and other bits and pieces. What does this huge Turing machine look like? Poundstone calculates that the whole construction, a self-reproducing machine incorporating a universal Turing machine, would be on the order of 10^{13} pixels.

> Displaying a 10^{13}-pixel pattern would require a video screen about 3 million pixels across at least. Assume the pixels are 1 millimeter square (which is very high resolution by the standards of home computers). Then the screen would have to be 3 kilometers (about two miles) across. It would have an area about six times that of Monaco.

> Perspective would shrink the pixels of a self-reproducing pattern to invisibility. If you got far enough away from the screen so that the entire pattern was comfortably in view, the pixels (and even the gliders, eaters and guns) would be too tiny to make out. A self-reproducing pattern would be a hazy glow, like a galaxy (*ibid.*, pp. 227–8).

Now, since the universal Turing machine can compute any computable function, it can play chess—simply by mimicking the program of any chess-playing computer you like. Suppose, then, that such an entity occupies the Life plane, playing chess against itself. Looking at the configuration of dots that accomplishes this marvel would almost certainly be unilluminating to anyone who had no clue that a configuration with such powers could exist. But from the perspective of one who had the hypothesis that this huge array of black dots was a chess-playing computer, enormously efficient ways of predicting the future of that configuration are made available. As a first step one can shift from an ontology of gliders and eaters to an ontology of symbols and machine states, and, adopting this higher design stance toward the configuration, predict its future *as* a Turing machine. As a second and still more efficient step, one can shift to an ontology of chess-board positions, possible chess moves, and the grounds for evaluating them; then, adopting the intentional stance toward the configuration, one can predict its future *as* a chess player performing intentional actions—making chess moves and trying to achieve checkmate. Once one has fixed on an interpretation scheme, permitting one to say which configurations of pixels count as which symbols (either, at the Turing machine level, the symbols '0' or '1', say, or at the intentional level, '*QxBch*' and the other symbols for chess moves), one can use the interpretation scheme to predict, for instance, that the next configuration to emerge from the galaxy will be such-and-such a glider stream (the symbols for '*RxQ*', say). There is risk involved in either case, because the chess program being run on the Turing machine may be far from perfectly rational, and, at a different level, debris may wander onto the scene and "break" the Turing machine configuration before it finishes the game.

In other words, real but (potentially) noisy patterns abound in such a configuration of the Life world, there for the picking up if only we are lucky or clever enough to hit on the right perspective. They are not ***visual*** patterns but, one might say, ***intellectual*** patterns. Squinting or twisting the page is not apt to help, while posing fanciful interpretations (or what W. V. Quine would call "analytical hypotheses") may uncover a goldmine. The opportunity confronting the observer of such a Life world is analogous to the opportunity confronting the cryptographer staring at a new patch of cipher text, or the opportunity confronting the Martian, peering through a telescope at the Superbowl Game. If the Martian hits on the intentional stance—or folk psychology—as the right level to look for pattern, shapes will readily emerge through the noise.

IV THE REALITY OF INTENTIONAL PATTERNS

The scale of compression when one adopts the intentional stance toward the two-dimensional chess-playing computer galaxy is stupendous: it is the difference between figuring out in your head what white's most likely (best) move is versus calculating the state of a few trillion pixels through a few hundred thousand generations. But the scale of the savings is really no greater in the Life world than in our own. Predicting that someone will duck if you throw a brick at him is easy from the folk-psychological stance; it is and will always be intractable if you have to trace the photons from brick to eyeball, the neurotransmitters from optic nerve to motor nerve, and so forth.

For such vast computational leverage one might be prepared to pay quite a steep price in errors, but in fact one belief that is shared by all of the representatives on the spectrum I am discussing is that "folk psychology" provides a description system that permits highly reliable prediction of human (and much nonhuman) behavior.[19] They differ in the explanations they offer of this predictive prowess, and the implications they see in it about "realism."

For Fodor, an industrial-strength Realist, beliefs and their kin would not be real unless the pattern dimly discernible from the perspective of folk psychology could also be discerned (more clearly, with less noise) as a pattern of structures in the brain. The pattern would have to be discernible from the different perspective provided by a properly tuned *syntactoscope* aimed at the purely formal (nonsemantic) features of Mentalese terms written in the brain. For Fodor, the pattern seen through the noise by everyday folk psychologists would tell us nothing about reality, unless it, and the noise, had the following sort of explanation: what we discern from the perspective of folk psychology is the net effect of two processes: an ulterior, hidden process wherein the pattern exists quite pure, overlaid, and partially obscured by various intervening sources of noise: performance errors, observation errors, and other more or less random obstructions. He might add that the interior belief-producing process was in this respect *just* like the process responsible for the creation of frames *A–F*. If you were permitted to peer behind the scenes at the program I devised to create the frames, you would see, clear as a bell, the perfect bar-code periodicity, with the noise thrown on afterward like so much salt and pepper.

---

[19] To see that the opposite poles share this view, see Fodor, *Psychosemantics* (Cambridge MIT, 1987), ch 1, "Introduction. the Persistence of the Attitudes"; and Paul Churchland, *Scientific Realism and the Plasticity of Mind* (New York. Cambridge, 1979), esp. p. 100: "For the P-theory [folk psychology] is in fact a marvelous intellectual achievement. It gives its possessor an explicit and systematic insight into the behaviour, verbal and otherwise, of some of the most complex agents in the environment, and its overall prowess in that respect remains unsurpassed by anything else our considerable theoretical efforts have produced"

This is often the explanation for the look of a data set in science, and Fodor may think that it is either the only explanation that can ever be given, or at any rate the only one that makes any sense of the success of folk psychology. But the rest of us disagree. As G. E. M. Anscombe[20] put it in her pioneering exploration of intentional explanation, "if Aristotle's account [of reasoning using the practical syllogism] were supposed to describe actual mental processes, it would in general be quite absurd. The interest of the account is that it describes an order which is there whenever actions are done with intentions . . ." (*ibid.*, p. 80).

But how *could* the order be there, so visible amidst the noise, if it were not the direct outline of a concrete orderly process in the background? Well, it *could* be there thanks to the statistical effect of very many concrete minutiae producing, as if by a hidden hand, an approximation of the "ideal" order. Philosophers have tended to ignore a variety of regularity intermediate between the regularities of planets and other objects "obeying" the laws of physics and the regularities of rule-following (that is, rule-*consulting*) systems.[21] These intermediate regularities are those which are preserved under selection pressure: the regularities dictated by principles of good design and hence homed in on by self-designing systems. That is, a "rule of thought" may be much more than a mere regularity; it may be a *wise* rule, a rule one would design a system by if one were a system designer, and hence a rule one would expect self-designing systems to "discover" in the course of settling into their patterns of activity. Such rules no more need be explicitly represented than do the principles of aerodynamics that are honored in the design of birds' wings.[22]

---

[20] *Intention* (New York: Blackwell, 1957).

[21] A notable early exception is Sellars, who discussed the importance of just this sort of regularity in "Some Reflections on Language Games," *Philosophy of Science*, XXI (1954)· 204–228. See especially the subsection of this classic paper, entitled "Pattern Governed and Rule Obeying Behavior," reprinted in Sellars's *Science, Perception and Reality*, pp. 324–7.

[22] Several interpreters of a draft of this article have supposed that the conclusion I am urging here is that beliefs (or their contents) are *epiphenomena* having no causal powers, but this is a misinterpretation traceable to a simplistic notion of causation If one finds a predictive pattern of the sort just described one has *ipso facto* discovered a causal power—a difference in the world that makes a subsequent

The contrast between these different sorts of pattern-generation processes can be illustrated. The frames in figure 1 were created by a hard-edged process (ten black, ten white, ten black, . . .) obscured by noise, while the frames in figure 4 were created by a process almost the reverse of that: the top frame shows a pattern created by a normal distribution of black dots around means at $x = 10, 30, 50, 70$, and 90 (rather like Mach bands or interference fringes); the middle and bottom frames were created by successive applications of a very simple contrast enhancer applied to the top frame: a vertical slit "window" three pixels high is thrown randomly onto the frame; the pixels in the window vote, and majority rules. This gradually removes the salt from the pepper and the pepper from the salt, creating "artifact" edges such as those discernible in the bottom frame. The effect would be more striking at a finer pixel scale, where the black merges imperceptibly through grays to white but I chose to keep the scale at the ten-pixel period of bar code. I do not mean to suggest that it is impossible to tell the patterns in figure 4 from the patterns in figure 1. Of course it is possible; for one thing, the process that produced the frames in figure 1 will almost always show edges at exactly 10, 20, 30, . . . and almost never at 9, 11, 19, 21, . . . while there is a higher probability of these "displaced" edges being created by the process of figure 4 (as a close inspection of figure 4 reveals). Fine tuning could of course reduce these probabilities, but that is not my point. My point is that *even if* the evidence is substantial that the discernible pattern is produced by one process rather than another, it can be rational to ignore those differences and use the simplest pattern description (e.g., *bar code*) as one's way of organizing the data.

Fodor and others have claimed that an interior language of thought is the best explanation of the hard edges visible in "propositional attitude psychology." Churchland and I have offered an alternative explanation of these edges, an explanation for which the

---

difference testable by standard empirical methods of variable manipulation. Consider the crowd-drawing power of a sign reading "Free Lunch" placed in the window of a restaurant, and compare its power in a restaurant in New York to its power in a restaurant in Tokyo. The intentional level is obviously the right level at which to predict and explain such causal powers; the sign more reliably produces a particular belief in one population of perceivers than in the other, and variations in the color of typography of the sign are not as predictive of variations in crowd-drawing power as are variations in (perceivable) meaning. The fact that the regularities on which these successful predictions are based are efficiently capturable (only) in intentional terms and are not derived from "covering laws" does not show that the regularities are not "causal"; it just shows that philosophers have often relied on pinched notions of causality derived from exclusive attention to a few examples drawn from physics and chemistry. Smith has pointed out to me that here I am echoing Aristotle's claim that his predecessors had ignored final causes.

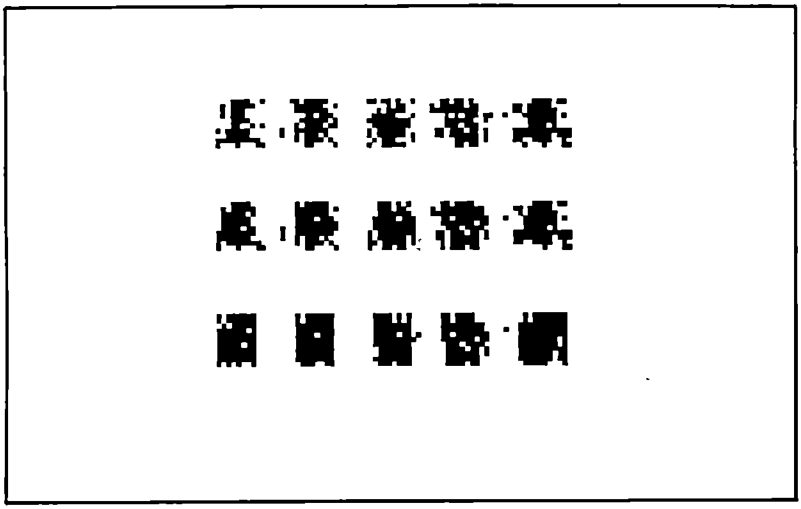


Figure 4

process that produced the frames in figure 4 is a fine visual metaphor. The process that produces the data of folk psychology, we claim, is one in which the multidimensional complexities of the underlying processes are projected *through linguistic behavior,* which creates an appearance of definiteness and precision, thanks to the discreteness of words.[23] As Churchland[24] puts it, a person's declarative utterance is a "one-dimensional *projection*—through the compound lens of Wernicke's and Broca's areas onto the idiosyncratic surface of the speaker's language—a one-dimensional projection of a four- or five-dimensional 'solid' that is an element in his true kinematic state" (*ibid.,* p. 85).

Fodor's industrial-strength Realism takes beliefs to be things in the head—just like cells and blood vessels and viruses. Davidson and I both like Churchland's alternative idea of propositional-attitude statements as indirect "measurements" of a reality diffused in the behavioral dispositions of the brain (and body).[25] We think beliefs are

[23] See my discussion of the distinction between beliefs and (linguistically infected) *opinions;* in *Brainstorms* (Montgomery, VT: Bradford, 1978), ch. 16, and in "The Illusions of Realism," in *The Intentional Stance,* pp. 110–6.

[24] "Eliminative Materialism and the Propositional Attitudes," this JOURNAL, LXXVIII, 2 (February 1981): 67–90, esp. p. 85.

[25] Churchland introduces the idea in *Scientific Realism and the Plasticity of Mind,* pp 100–7. My adoption of the idea was in "Beyond Belief," in A. Woodfield, ed., *Thought and Object* (New York: Oxford, 1982), repr. as ch. 5 of *The Intentional Stance.* Davidson's guarded approval is expressed in "What is Present to the Mind?" read at the Sociedad Filosófica Ibero Americana meeting in Buenos Aires, 1989.

quite real enough to call real just so long as belief talk measures these complex behavior-disposing organs as predictively as it does. What do we disagree about? As John Haugeland[26] has pointed out, Davidson is more of a realist than I am, and I have recently tracked down the source of this disagreement to a difference of opinion we have about the status of Quine's principle of indeterminacy of translation, which we both accept.

For Davidson, the principle is not the shocker it is often taken to be; in fact, it is well-nigh trivial—the two different translation manuals between which no fact of the matter decides are like two different scales for measuring temperature.

> We know there is no contradiction between the temperature of the air being 32° fahrenheit and 0° celsius; there is nothing in this 'relativism' to show that the properties being measured are not 'real'. Curiously, though, this conclusion has repeatedly been drawn. . . . Yet in the light of the considerations put forward here, this comes to no more than the recognition that more than one set of one person's utterances might be equally successful in capturing the contents of someone else's thoughts or speech. Just as numbers can capture all the empirically significant relations among weights or temperatures in infinitely many ways, so one person's utterances can capture all the significant features of another person's thoughts and speech in different ways. This fact does not challenge the 'reality' of the attitudes or meanings thus reported.[27]

On Davidson's view, no substantive disagreements emerge from a comparison of the two description schemes, and so they can quite properly be viewed as competing descriptions of the same reality.

I think this is a flawed analogy. A better one is provided by the example of "rival" descriptions of patterns-with-noise. Consider two rival intentional interpretations of a single individual; they agree on the general shape of this individual's collection of beliefs (and desires, etc), but because of their different idealizations of the pattern, they do not agree point-for-point. Recalling a famous analogy of Quine's[28] and extending it beyond radical translation to radical interpretation (as Davidson and I both wish to do), we get the image in Figure 5.

To the left we see Brown's intentional interpretation of Ella; to the right, Jones's interpretation. Since these are intentional interpreta-

---

[26] See the discussion of Haugeland's views in the last chapter of *The Intentional Stance*, "Mid-Term Examination Compare and Contrast," pp. 348–9

[27] Davidson, "What is Present to the Mind?" (ms.), p. 10.

[28] "Different persons growing up in the same language are like different bushes trimmed and trained to take the shape of identical elephants. The anatomical details of twigs and branches will fulfill the elephantine form differently from bush to bush, but the overall outward results are the same." *Word and Object* (Cambridge· MIT, 1960), p. 8.

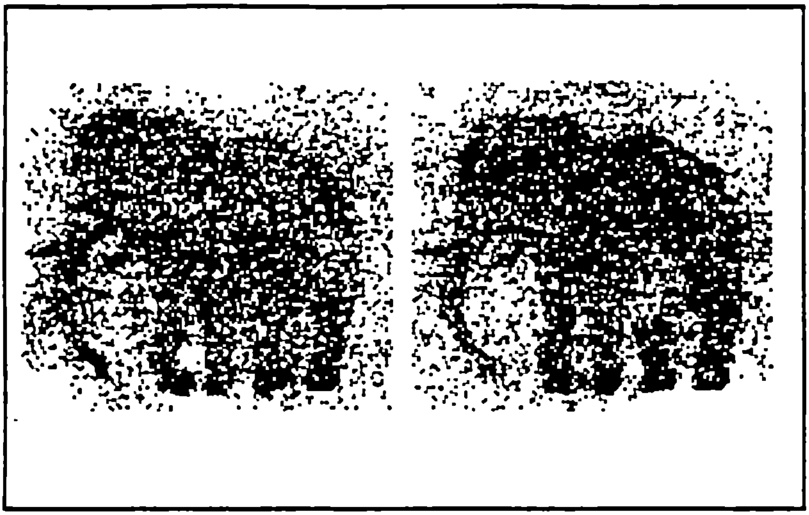


Figure 5

tions, the pixels or data points represent beliefs and so forth, not (for instance) bits of bodily motion or organs or cells or atoms, and since these are rival intentional interpretations of a single individual, the patterns discerned are not statistical averages (e.g., "Democrats tend to favor welfare programs") but personal cognitive idiosyncracies (e.g., "She thinks she should get her queen out early"). Some of the patterns may indeed be simple observed periodicities (e.g., "Ella wants to talk about football on Mondays") but we are to understand the pattern to be what Anscombe called the "order which is there" in the rational coherence of a person's set of beliefs, desires, and intentions.

Notice that here the disagreements can be substantial—at least before the fact: when Brown and Jones make a series of predictive bets, they will not always make the same bet. They may *often* disagree on what, according to their chosen pattern, will happen next. To take a dramatic case, Brown may predict that Ella will decide to kill herself; Jones may disagree. This is not a trivial disagreement of prediction, and in principle this momentous difference may emerge in spite of the overall consonance of the two interpretations.

Suppose, then, that Brown and Jones make a series of predictions of Ella's behavior, based on their rival interpretations. Consider the different categories that compose their track records. First, there are the occasions where they agree and are right. Both systems look good from the vantage point of these successes. Second, there are the occasions where they agree and are wrong. Both chalk it up to noise, take their budgeted loss and move on to the next case. But there will also be the occasions where they disagree, where their systems make different predictions, and in these cases sometimes (but not always) one will win and the other lose. (In the real world, predictions are not always from among binary alternatives, so in many cases they will disagree and both be wrong.) When one wins and the other loses, it will look to the myopic observer as if one "theory" has scored a serious point against the other, but when one recognizes the possibility that both may chalk up such victories, and that there may be no pattern in the victories which permits either one to improve his theory by making adjustments, one sees that local triumphs may be insufficient to provide any ground in reality for declaring one account a closer approximation of the truth.

Now, some might think this situation is *always* unstable; eventually one interpretation is bound to ramify better to new cases, or be deducible from some larger scheme covering other data, etc. That might be true in many cases, but—and this, I think, is the central point of Quine's indeterminacy thesis—it need not be true in all. *If* the strategy of intentional-stance description is, as Quine says, a "dramatic idiom" in which there is ineliminable use of idealization, and if Fodor's industrial-strength Realism is thus not the correct explanation of the reliable "visibility" of the pattern, such radical indeterminacy is a genuine and stable possibility.

This indeterminacy will be most striking in such cases as the imagined disagreement over Ella's suicidal mindset. If Ella does kill herself, is Brown shown to have clearly had the better intentional interpretation? Not necessarily. When Jones chalks up his scheme's failure in this instance to a bit of noise, this is no more ad hoc or unprincipled than the occasions when Brown was wrong about whether Ella would order the steak not the lobster, and chalked those misses up to noise. This is not at all to say that an interpretation can never be shown to be just wrong; there is plenty of leverage within the principles of intentional interpretation to refute particular hypotheses—for instance, by forcing their defense down the path of Pickwickian explosion ("You see, she didn't believe the gun was loaded because she thought that those bullet-shaped things were chocolates wrapped in foil, which was just a fantasy that occurred to her because . . . ."). It *is* to say that there could be two interpretation schemes that were reliable and compact predictors over the long run, but that nevertheless disagreed on crucial cases.

It might seem that in a case as momentous as Ella's intention to kill herself, a closer examination of the details just prior to the fatal moment (if not at an earlier stage) would have to provide additional support for Brown's interpretation at the expense of Jones's interpretation. After all, there would be at least a few seconds—or a few hundred milliseconds—during which Ella's decision to pull the trigger got implemented, and during that brief period, at least, the evidence would swing sharply in favor of Brown's interpretation. That is no doubt true, and it is *perhaps* true that had one gone into enough detail earlier, all this last-second detail could have been predicted—but to have gone into *those* details earlier would have been to drop down from the intentional stance to the design or physical stances. From the intentional stance, these determining considerations would have been invisible to both Brown and Jones, who were both prepared to smear over such details as noise in the interests of more practical prediction. Both interpreters concede that they will make false predictions, and moreover, that when they make false predictions there are apt to be harbingers of misprediction in the moments during which the *dénouement* unfolds. Such a brief swing does not constitute refutation of the interpretation, any more than the upcoming misprediction of behavior does.

How, then, does this make me less of a realist than Davidson? I see that there could be two different systems of belief attribution to an individual which differed *substantially* in what they attributed—even in yielding substantially different predictions of the individual's future behavior—and yet where no deeper fact of the matter could establish that one was a description of the individual's *real* beliefs and the other not. In other words, there could be two different, but equally real, patterns discernible in the noisy world. The rival theorists would not even agree on which parts of the world were pattern and which were noise, and yet nothing deeper would settle the issue.[29] The choice of a pattern would indeed be up to the observer, a matter to be decided on idiosyncratic pragmatic grounds. I myself do not see any feature of Davidson's position that would be a serious obstacle to his shifting analogies and agreeing with me. But then he would want to grant that indeterminacy is not such a trivial matter after all.[30]

What then is Rorty's view on these issues? Rorty wants to deny that any brand of "realism" could *explain* the (apparent?) success of the intentional stance. But since we have already joined Fine and set aside the "metaphysical" problem of realism, Rorty's reminding us of this only postpones the issue. Even someone who has transcended the scheme/content distinction and has seen the futility of correspondence theories of truth must accept the fact that *within* the natural ontological attitude we sometimes explain success by correspondence: one does better navigating off the coast of Maine when one uses an up-to-date nautical chart than one does when one uses a road map of Kansas. Why? Because the former accurately represents the hazards, markers, depths, and coastlines of the Maine coast, and the latter does not. Now why does one do better navigating the shoals of interpersonal relations using folk psychology than using astrology? Rorty might hold that the predictive "success" we folk-psychology players relish is itself an artifact, a mutual agreement engendered by the egging-on or consensual support we who play this game provide each other. He would grant that the game has no rivals in popularity, due—in the opinion of the players—to the power it gives them to understand and anticipate the animate world. But he would refuse to endorse this opinion. How, then, would he distinguish this popularity from the popularity among a smaller coterie of astrology?[31] It is undeniable that astrology provides its adherents with a highly articulated system of patterns that they *think* they see in the events of the world. The difference, however, is that no one has ever been able to get rich by betting on the patterns, but only by selling the patterns to others.

---

[29] Cf "The Abilities of Men and Machines," in *Brainstorms*, where I discuss two people who agree exactly on the future behavior of some artifact, but impose different Turing-machine interpretations of it. On both interpretations, the machine occasionally "makes errors" but the two interpreters disagree about which cases are the errors. (They disagree about which features of the object's behavior count as signal and which as noise.) Which Turing machine is it really? This question has no answer.

[30] Andrej Zabludowski seems to me to have overlooked this version of indeterminacy in "On Quine's Indeterminacy Doctrine," *Philosophical Review*, XCVIII (1989): 35–64.

Rorty would have to claim that this is not a significant difference; the rest of us, however, find abundant evidence that our allegiance to folk psychology as a predictive tool can be defended in coldly objective terms. We agree that there is a real pattern being described by the terms of folk psychology. What divides the rest of us is the nature of the pattern, and the ontological implications of that nature.

Let us finally consider Churchland's eliminative materialism from this vantage point. As already pointed out, he is second to none in his appreciation of the power, to date, of the intentional stance as a strategy of prediction. Why does he think that it is nevertheless doomed to the trash heap? Because he anticipates that neuroscience will eventually—perhaps even soon—discover a pattern that is so clearly superior to the noisy pattern of folk psychology that everyone will readily abandon the former for the latter (except, perhaps, in the rough-and-tumble of daily life). This might happen, I suppose. But

---

[31] Cf. my comparison of "the astrological stance" to the intentional stance, *The Intentional Stance*, p 16.

Churchland here is only playing a hunch, a hunch that should not be seen to gain plausibility from reflections on the irresistible forward march of science. For it is not enough for Churchland to suppose that in principle, neuroscientific levels of description will explain more of the variance, predict more of the "noise" that bedevils higher levels. This is, of course, bound to be true in the limit—if we descend all the way to the neurophysiological "bit map." But as we have seen, the trade-off between ease of use and immunity from error for such a cumbersome system may make it profoundly unattractive.[32] If the "pattern" is scarcely an improvement over the bit map, talk of eliminative materialism will fall on deaf ears—just as it does when radical eliminativists urge us to abandon our ontological commitments to tables and chairs. A truly general-purpose, robust system of pattern description more valuable than the intentional stance is not an impossibility, but anyone who wants to bet on it might care to talk to me about the odds they will take.

What does all this show? Not that Fodor's industrial-strength Realism must be false, and not that Churchland's eliminative materialism must be false, but just that both views are gratuitously strong forms of materialism—presumptive theses way out in front of the empirical support they require. Rorty's view errs in the opposite direction, ignoring the impressive empirical track record that distinguishes the intentional stance from the astrological stance. Davidson's intermediate position, like mine, ties reality to the brute existence of pattern, but Davidson has overlooked the possibility of two or more *conflicting* patterns being superimposed on the same data—a more radical indeterminacy of translation than he had supposed possible. Now, once again, is the view I am defending here a sort of instrumentalism or a sort of realism? I think that the view itself is clearer than either of the labels, so I shall leave that question to anyone who stills find illumination in them.

DANIEL C. DENNETT

Tufts University

---

[32] As I have put it, physical-stance predictions trump design-stance predictions, which trump intentional-stance predictions—but one pays for the power with a loss of portability and a (usually unbearable) computational cost.

BOOK REVIEWS

The Nature of Mental Things. ARTHUR W. COLLINS. Notre Dame: Notre Dame University Press, 1987. xviii, 178 p. Cloth $22.95.

Is philosophy prone to fashion? And if so, can we think of a philosophical position like a dress that we can store in the attic to await the time when the current fashion makes it acceptable to wear again? Reading *The Nature of Mental Things* puts one in mind of a good rummage around in the attic. The things one finds are not so very old—just a generation or so. Some of what one finds is a little dusty, and some misses the point of what has come since, but much of what one finds makes one reflect upon the life that went before; if one is sufficiently open, it can teach one much about the present.

In his recent book, Arthur W. Collins is concerned with the nature of mind. He writes globally of "the mind," but his real concern is with beliefs, desires, and visual experiences. He aims to show that "there is less 'to the mind' than we philosophers have supposed" (43). He criticizes at length identity theorists who, he suggests, may be either materialists or dualists. The functionalist, who claims that mental states are realized in states of a body or of a soul, is attacked by association. Those philosophers who, in accounting for propositional attitudes, invoke the notion of representation, of languages of thought, of information processing, or of computation are accused of simply turning out variations on a mistaken theme. The mistake is in the assumption that the mind is an inner thing. Heavily influenced by Ludwig Wittgenstein, Collins aims to diagnose how the mistake comes to be made and to dispel the illusion.

Collins insists that common-sense thinking about the mind is adulterated by theory, and he attempts to pick apart the strands. Here are some pretheoretical (and acceptable) reflections: I have beliefs and so do others; some beliefs are true and some are false; the mind is private—that is, most people know about their own mental states much of the time and can tell others about these states with a certain authority; there are states of consciousness—for example, people who see have visual experiences. Here is what philosophical theory (mistakenly) adds to these basic reflections: the mind is an object of privileged access via infallible introspective consciousness; beliefs are neural states or functional states or states of an immortal soul; there is a phenomenon of consciousness to which we attempt to refer by

0022-362X/91/8801/52–6

such expressions as 'immediate experience' or 'inner goings on'. What theory adds to common sense, and where our thinking about the mind goes awry, is in the idea that "beliefs, desires, and other mental things are inner states or processes" (2). Despite leaning heavily on common sense, Collins claims not to rest his case on intuition but on the fact that inner-state theories cannot do the job required of them. It is unclear whether Collins thinks we need a better theory or whether he thinks that the real problem is that we attempt to theorize at all.

Philosophical thinking about consciousness provides a clear illustration of how we go wrong. It starts with the idea of the phenomena of consciousness. It is already on the wrong footing because, claims Collins, phenomena are what a theory must save (64). Its next step is to convert the phenomena into objects of consciousness. Finally, the phenomena, the objects, are identified with either a Cartesian immaterial soul or a material body. (Here we see that Cartesian dualism and materialism are just two sides of the same coin—a point much emphasized by Collins.) The point of the identification is to give an account of the objects of consciousness. But, argues Collins, an account of the objects of consciousness is one thing, an account of consciousness is another. Collins suggests that inner-state theories may be "rescued" by the following move: collapse the object of consciousness into consciousness itself. Thus we are presented with the idea that the objects of consciousness are luminous, they are intrinsically conscious. But now we need some account of this luminosity and, claims Collins, none is forthcoming. Collins concludes that inner-state theories cannot do the job they are required to do. But perhaps we are only entitled to conclude that, as they now stand, inner-state theories do not do the job required.

According to Collins, once we adopt this theory and latch onto the idea of an inner reality, we are doomed to shuffle back and forth between dualism and materialism. We must exorcise the theory. The concept of mind is "outer-oriented" (175). In this respect, analytical behaviorism got things right. But are we not now in danger of shuffling back and forth between the inner and the outer? Having turned our backs on the idea of the mind as an *inner* reality, are we now to swallow the idea (the theory?) of the mind as an *outer* reality? Collins is much less forthcoming here (although his tolerance of analytical behaviorism may be indicative). His attack on eliminativism as a "lunatic philosophy of mind" leads one to conclude that it is not ~~the~~ reality of the mental so much as its innerness that is being rejected.

The suggestion that the mind may not be wholly inner but *also* inner is not considered. One is left wondering whether things would not have looked different if Collins had chosen to concentrate less on the case of visual experience and instead had considered sensations like pain. (A curious footnote appears on page 20 where Collins claims to agree with Wittgenstein that the mental is not an inner state or process yet to disagree with the same over the case of pain. Pains, claims Collins, *are* inner realities.) And even if one does look to visual experience, it is unclear why "his seeing of the wax" is a particularly problematic accompaniment of "he sees the wax." Collins does not accuse these nominalizations of tempting us to repeat after Franz Brentano that inner perception is the only real perception. Direct realism is compatible with the acceptance of introspective awareness. Nor does introspective awareness demand infallibility. One is left to understand that the real problem here is that nominalization tempts one to inner objects, and inner objects tempt one to identifications that leave all that matters unexplained. But, once again, this only shows that the theory has more work to do, not that it is doomed to fail.

Collins discusses Thomas Nagel's work as a case in point: talk of the "subjectivity of experience," of "what it is like," and of "points of view" all succumbs to the temptation of the inner. The fact that Nagel rejects all current forms of reductive materialism but holds out the hope of some possible future reduction only serves to strengthen Collins's point. But there is an aspect of Nagel's work which Collins overlooks: the mental according to Nagel is not just subjective, it is not objective. That is to say, it cannot be understood from an impersonal or detached perspective. (Latterly, Nagel has distinguished physical from mental objectivity—the mind has a *kind* of objectivity, mental objectivity. This is an acknowledgment that there are minds other than our own, but it is also an acknowledgment that, from the outside, we can achieve only an incomplete conception of these minds.) Once we accept that the mind is not objective, cannot be so understood, we should be less inclined simply to move from the mental as inner object to the mental as outer object. It may not be the idea of the innerness of the mental which causes problems so much as that that which is subjective is mistaken for that which is objective. And once we appreciate this, it becomes clear that Cartesian dualism is as much of a mistake as current forms of materialism. The latter is a point much emphasized by Nagel—although Collins writes as if this parallel were overlooked by Nagel (cf. 98, 99).

For a proper rejection of mind-body identifications, Collins turns from Nagel's work to Saul Kripke's. Collins wants to push Kripke's conclusion one step further: once we find that theoretical identifications of visual experience (Collins's example) fail, we ought to reject the very assumption on which they rest: the idea that visual experiences are inner items that call for an identification. We ought to reject the phenomena of visual experience. Perhaps Collins could be clearer about the extent of his divergence from Kripke. The latter after all makes much of a "Cartesian intuition." And there is another difference. Kripke aims to shift a certain burden onto those who claim an identity: explain the intuition of contingency showing it to be an illusion or a mistake. Kripke considers and rejects an explanation that requires that we distinguish mental phenomena and our experiences of them. (The rejection of this explanation does seem even more obvious for Collins's case of visual experience than for Kripke's case of pain.) Collins follows Kripke here, yet Collins concludes that "all theoretical identifications of the form 'visual experiences are . . .' fail" (117). Kripke merely concludes that "the usual moves and analogies are not available to solve the problems of the identity theorist," and he adds that this "is, of course, no proof that no moves are available."[1] It remains open to the identity theorist, then, to come up with some other explanation of the intuition. And some identity theorists believe they have such an alternative explanation—albeit one that relies on making the move from type- to token-identity theories. Token-identity theories are only briefly discussed by Collins—presumably because all identity theories share the same mistaken assumption of inner phenomena.

Collins suggests that the identity theories' real interest is in token mental events and that the move from these to types depends upon the discovery of psychophysical laws. He cites Donald Davidson as an example of someone whose pessimism over such a discovery leaves him embracing only a token-identity theory. Collins's diagnosis, however, does leave us wondering why there is such an interest in token identities. He would no doubt reply that it is part of a mistaken conception of mind, but I would suggest a different answer: the apparent causal interaction between the mental and the physical.

Admitting causal relations here will require that we salvage some part of the idea of the mental as inner—and this, Collins will argue, takes us back to the mistaken theory. But it is a real question whether these relations *are* part of the theory or whether they may be invoked on the side of our pretheoretical thinking about mind. Collins suggests that it is positivist thinking about *explanation* which gives encouragement to identity theories of mind: if all explanation is causal explanation, then beliefs and desires are the obvious candidates to figure in the explanation of actions; these beliefs and desires are then taken to be inner realities that must be physical realities if they are to be causally effective in a physical world. In order to avoid this conception of the mind, we need to shift away from a causal account of reason giving. Collins suggests that teleological explanation should serve as the model for action explanation. This form of explanation "renounces reference to antecedent causes in favour of reference to the outcome" (144). He devotes the final two chapters of his book to a defense of the position that action explanation is not causal explanation, but he does not consider whether a causal condition is necessary for action. The argument in favor of the causal condition is as follows: a man may have a reason to act and he may act, without its being the case that that reason was the reason for that act. Collins may reply that such an argument relies on a prior acceptance of the misguided conception of mind. But one could rejoin by pointing out that it is our conception of action that demands the causal condition.

---

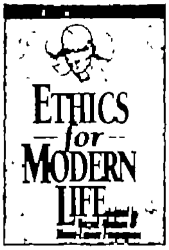
[1] *Naming and Necessity* (Cambridge Harvard, 1980), p 155

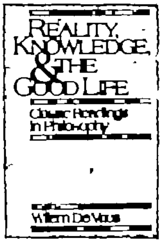
The real strength of Collins's work is to be found in the wide perspective he offers. He forces us to re-examine so much of what has come to be taken for granted in some philosophical circles. It seems to me just right that what needs our attention these days is an examination of the conception of mind with which we are working. In this, Collins's work is exemplary.

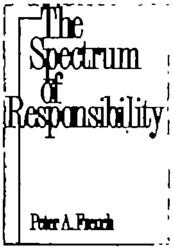
ANITA AVRAMIDES

St. Hilda's College, Oxford University

Foundations for Thought




REALITY, KNOWLEDGE & THE GOOD LIFE
Classic Readings in Philosophy
Willem DeVries



Ethics for Modern Life *Fourth Edition*

Edited by RAZIEL ABELSON, New York University
MARIE-LOUISE FRIQUEGNON, William Paterson College

A well balanced introduction to both theoretical and applied ethics that first examines the standards of moral right and wrong and the grounds of moral responsibility and then applies these basic philosophic principles to modern social issues. The chapter on environmental ethics is new to this edition, and the chapter on suicide has been reinstated from the Second Edition.

Paperbound. 480 pages. Publication: January 1991

Reality, Knowledge, and the Good Life

A HISTORICAL INTRODUCTION TO PHILOSOPHY
Edited by WILLEM A. DEVRIES, University of New Hampshire

A text-reader containing 31 selections from 18 major philosophers. The text presents three major themes in the history of philosophy - reality, knowledge, and the good life - and shows how the themes have evolved over time in the writings of important philosophers from the early Greeks to the 20th century.

Paperbound. 704 pages (probable). Publication: February 1991

The Spectrum of Responsibility

PETER A. FRENCH, Trinity University

A brief anthology that brings together differing viewpoints on the concept of moral responsibility -- specifically, the responsibility of individuals and groups to each other and to society as a whole. Responsibility is examined at both the collective and personal level, offering students the opportunity to see moral roles in the business or professional world.

Paperbound. 236 pages. Publication: January 1991

ST. MARTIN'S PRESS
College Division, Department JR
175 Fifth Avenue, New York, N.Y. 10010

To request an examination copy of any of the titles listed in this ad, please call us at our toll-free number, 1-800-446-8923, or write us on your college letterhead specifying course title, present text, and enrollment. Send your request to the above address.

Frontiers of Philosophy

The Philosopher's Joke

Richard A. Watson

In this unorthodox volume, Richard A. Watson makes merry with the styles of philosophy fashionable today and in each of the last four decades, in such essays as "The Seducer and the Seduced," "Ape Dreams," and "A Pig's Tail." Although the essays embrace fundamental philosophical arguments, they are written with an attention to style not found in conventional philosophical texts. Such concern with matters of style, while it is the most stimulating and original aspect of the book, is likely to be what most unsettles anyone who expects a book of philosophy to have a tightly knit logical structure The essays *are* droll; nonetheless, each has philosophical content that is examined seriously.

124 pages • Paper $14 95

Doing Philosophy Historically

edited by Peter H. Hare

Can original philosophy be done while simultaneously engaging in the history of philosophy? The contributors to this work affirm that historical philosophy is a valuable enterprise; they include Todd L Adams, Lilli Alanen, José Benardete, Jonathan Bennett, John I. Biro, Phillip Cummins, Georges Dicker, Daniel A. Dombrowski, Daniel Garber, Josiah B. Gould, Jorge J. E. Gracia, Daniel W. Graham, Charles Griswold, James Lawler, Rudolf Luthe, Edward H. Madden, George Mavrodes, Gerald E. Myers, Jonathan Rée, Frithjof Rodi, Kenneth L. Schmitz, Vladimir Shtinov, David G. Stern, Robert Turnbull, James Van Cleve, Frederick P Van De Pitte, Henry Veatch, and Richard A. Watson

"Exciting . . .this book does not just present arguments that original philosophy can be done historically; it actually is a case itself of doing philosophy historically " —Choice

362 pages • Cloth $36 95

Animal Consciousness

Daisie Radner and Michael Radner

The authors argue that the usual reasons given to undermine the validity of animal consciousness are ill-founded, and explore the ethical implications of accepting nonhuman cognition.

"Well constructed and adroitly argued." —Nature

263 pages • Cloth $39 95

But Is It Science?

The Philosophical Question in the Creation/Evolution Controversy

edited by Michael Ruse

"Exquisite philosophical, scientific, and legal detail "

—Science Books & Films

406 pages • Cloth $38 95

Naturalism and Rationality

edited and with an introduction by Newton Garver and Peter H Hare

"A lasting source of stimulation for contemporary philosophy "

—Review Journal of Philosophy & Social Science

289 pages • Cloth $31 95

Latin American Philosophy in the Twentieth Century

Man, Values, and the Search for Philosophical Identity

edited by Jorge J. E Gracia

"Quite readable . . . clearheaded "

—Latin America in Books

269 pages • Paper $19 95

At better bookstores, or order directly from

PROMETHEUS BOOKS

Call toll free (800) 421-0351 • 59 John Glenn Drive • Amherst, NY 14228-2197

Add $2.50 p&h for first book, $1 25 for each additional book ($7 50 maximum) NY State residents add sales tax. #4003

Now published—the final volume in The Collected Works of John Dewey

"This splendid publishing venture by Southern Illinois University Press should serve less as a tombstone for John Dewey than as a stimulus for a new flourishing of American philosophy."
—*Philosophy and Phenomenological Research*

THE LATER WORKS OF JOHN DEWEY, 1925 -1953
Volume 17: 1885-1953
Miscellaneous Writings

Edited by JO ANN BOYDSTON, with an Introduction by SIDNEY HOOK
This final textual volume of The Collected Works contains Dewey's writings discovered after publication of the appropriate volume and spans most of Dewey's publishing life. Of the 83 items contained here, 24 of them have not been published previously. (1661-7) $50.00

Also of interest...

LECTURES ON ETHICS, 1900 -1901

JOHN DEWEY, edited and with an introduction by DONALD F. KOCH
Koch argues that these lectures offer the best systematic, overall introduction to Dewey's approach to moral philosophy and are the only account showing the unity of his views in nearly all phases of ethical inquiry. (1663-3) $55.00

JOHN CRAIGE'S *MATHEMATICAL PRINCIPLES OF CHRISTIAN THEOLOGY*

RICHARD NASH
First published in Latin in 1699, John Craige's *Theology* represents a rare early attempt to introduce mathematical reasoning into moral and theological dispute. (*Journal of the History of Philosophy* Monograph Series)
(1662-5) $15.95 paper

Add $2.00 when ordering by mail. VISA and MasterCard accepted.
Please include sales tax on all orders shipped to Illinois addresses.

SOUTHERN ILLINOIS UNIVERSITY PRESS
Dept. JP11, P.O. Box 3697, Carbondale, IL 62902-3697

Oxford

Reason and Religious Belief

An Introduction to the Philosophy of Religion

MICHAEL PETERSON, WILLIAM HASKER, BRUCE REICHENBACH, and DAVID BASINGER

"Comprehensive, well-written, up-to-date.... An excellent text for students approaching the subject for the first time." —Robert Audi, *University of Nebraska*. This text covers standard topics—religious experience, faith and reason, theistic arguments, the problem of evil, religious language, miracles and life after death—as well as new topics that have been widely discussed in the last fifteen years, including Reformed epistemology, the philosophical analysis of theological doctrine, and the Kalam cosmological argument.

1990 304 pp. paper $17.95

What is Justice?

Classic and Contemporary Readings

Edited by ROBERT C. SOLOMON and MARK C. MURPHY

This uniquely comprehensive anthology brings together substantial excerpts from Homer, Plato, Aristotle, the Bible, and the Quran, as well as Hobbes, Rousseau, Kant, Hegel, Engels, Rawls, Nozick, Walzer, and Nagel.

1990 394 pp. paper $14.95 cloth $38.00

The Nature of Mind

Edited by DAVID M. ROSENTHAL

This anthology provides a framework for understanding mental functioning, with works of Descartes, Locke, Reid, Fodor, Dennett, Nagel, Putnam, Davidson, Searle, Ryle, Strawson, Burge, Chisholm, Rorty, and Sellars.

February 1991 672 pp. paper $19.95 cloth $49.95

Pragmatics

A Reader

Edited by STEVEN DAVIS

Drawing on the work of philosophers, linguists, and psychologists, this text includes seminal papers by some of the most important recent writers on pragmatics, among them H.P. Grice, J.R. Searle, Saul Kripke, David Kaplan, Deirdre Wilson, and Dan Sperber.

January 1991 576 pp. paper $29.95

The Philosophy of Language

Second Edition

Edited by A.P. MARTINICH

"An excellent and classical collection. Indispensable."
—Richard N. Burnor, *Towson State University*

1990 544 pp. cloth $29 95

The Inner Citadel

Essays on Individual Autonomy

Edited by JOHN CHRISTMAN

This anthology includes essays by such eminent philosophers as Gerald Dworkin, Joel Feinberg, Harry Frankfurt, and David A.J. Richards.

1989 288 pp. paper $15.95 cloth $36.00

Contemporary Moral Controversies in Business
Edited by A. PABLO IANNONE
"Simply the best, most comprehensive anthology on business ethics available.... A very valuable teaching tool." —Frank K Fair, *Sam Houston State University*
1989 608 pp.; 3 illus. paper $21.00

Computers, Ethics, and Society
Edited by M. DAVID ERMANN, MARY B. WILLIAMS, and CLAUDIO GUTIERREZ
"Quite good.... Will serve a critical need of current writing on this topic." —R. Waldo Roth, *Taylor University*
1990 384 pp.; 1 illus., 3 maps paper $16.95

Critical Thinking
An Introduction to Reasoning
FRANCIS WATANABE DAUER
"An excellent choice for an honors course in informal logic/critical thinking." —Robert B. Scott, *William Woods College*
1989 560 pp.; 23 illus. cloth $34.50
Teacher's Manual 1989 paper 400 pp.; illus.
complimentary to instructors

An Anthology of Western Marxism
From Lukács and Gramsci to Socialist-Feminism
Edited by ROGER S. GOTTLIEB
"Excellent text.... The inclusion of feminist thought makes it doubly useful." —Susan Hunter, *West Virginia University.*
1989 400 pp. paper $15.95

What Does It All Mean?
A Very Short Introduction to Philosophy
THOMAS NAGEL
"An excellent introduction to the study of philosophy for persons with no philosophical background at all." —Robert M. Fowler, *University of Colorado*
1987 106 pp. paper $7.95 cloth $16.95

The View From Nowhere
THOMAS NAGEL
"Remarkable." —P.F. Strawson, *The New Republic*
1986 (paper 1989) 256 pp. paper $10.95 cloth $28.00

Seven Theories of Human Nature
Second Edition
LESLIE STEVENSON
"An excellent basic text." —Luise H. Morton, *Ohio State University*
1987 (paper 1988) 160 pp. paper $9.95 cloth $21.95

Prices and publication dates are subject to change.
To request an examination copy, write on school letterhead giving full course information, including course name, level, expected enrollment, and your decision deadline, to:
College Humanities & Social Sciences Marketing Department

Oxford University Press
200 Madison Avenue • New York, NY 10016

Oxford

Kant's Transcendental Psychology

PATRICIA KITCHER, *University of California, San Diego*

"Kitcher's work here concerns some of the most difficult and central aspects of the work of one of our greatest philosophers, and it comes at that work from a unique and very interesting perspective. She asks the kinds of questions that students and readers of Kant very much want to see addressed, and she does this in an exceptionally lively and knowledgeable way—for everyone, there is always something in her work to be learned from." —Karl Ameriks, *University of Notre Dame*

1990 312 pp. $35.00

Kant's Model of the Mind

A New Interpretation of Transcendental Idealism

WAYNE WAXMAN, *New School for Social Research, New York*

This book argues that Kant's transcendental idealism has been misinterpreted: it denies not simply the super-*sensory* reality of space, time, and appearances, but their reality outside *imagination* as well. Waxman shows this interpretation to be essential to the Transcendental Deduction, the affirmation of things in themselves, and the attempt to surmount Hume's scepticism.

1990 320 pp. $39.95

Freedom Within Reason

SUSAN WOLF, *Johns Hopkins University*

Offering a new view about the conditions of freedom and responsibility, Wolf here makes a case for freedom as the ability to act in accordance with one's values and the ability to form one's values in light of an appreciation of the True and the Good.

1990 176 pp. $24.95

The Dilemma of Freedom and Foreknowledge

LINDA ZAGZEBSKI, *Loyola Marymount University, Los Angeles*

This original analysis examines the three leading traditional solutions to the dilemma of divine foreknowledge and human free will—those arising from Boethius, Ockham, and Molina. Though all three solutions are rejected in their best-known forms, three new solutions are proposed, and Zagzebski concludes that divine foreknowledge is compatible with human freedom.

1990 256 pp. $29.95

Liberal Virtues

Citizenship, Virtue, and Community in Liberal Constitutionalism

STEPHEN MACEDO, *Harvard University*

"One of the most striking features of *Liberal Virtues* is the way that it unites the perspective of the ordinary citizen with the insights of academic political philosophy. The effect is to deepen citizen understanding while at the same time giving a much needed lesson in common sense to academics. This is a remarkable achievement." —Jeffrey K. Tulis, *University of Texas at Austin*

1990 320 pp. $29.95

Plato's Parmenides

CONSTANCE C. MEINWALD, *University of Illinois at Chicago*

"Simply stunning. (Meinwald's) interpretation of a difficult and crucial Platonic text is not just to be placed among others as one more excellent competitor. It supersedes the others." —Sandra Peterson, *University of Minnesota*. Meinwald shows that to understand Plato's response in the dialogue, his distinction between two kinds of predication needs to be recognized.

1990 224 pp. $29.95

Prices and publication dates are subject to change and good only in the U.S.
To order, send check or money order to: Department MG
To order by phone using major credit cards,
please call 1-800-334-4249 ext. 7106

Oxford University Press
200 Madison Avenue • New York, NY 10016

THE JOURNAL OF PHILOSOPHY

Philosophy)

VOLUME LXXXVIII, NUMBER 2
FEBRUARY 1991

page
57 *Irrational Actions* Rosalind Hursthouse
69 *Substance: Prolegomena to a Realist Theory of Identity* Michael Ayers
91 *The Indeterminacy Thesis Reformulated* Ken Gemes

BOOK REVIEWS
109 Steven M. Nadler: *Arnauld and the Cartesian Philosophy of Ideas* John Yolton

112 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY
FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us

© Copyright 1991 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER Second class postage paid at New York, NY, and additional mailing offices Send address changes to *Journal of Philosophy*, 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NO. 2, FEBRUARY 1991

ARATIONAL ACTIONS*

It is often said that there is some special irrationality involved in wreaking damage or violence on inanimate objects that have angered one, and, correspondingly, something rational about striking people or animals in anger. The explanation of this seems obvious, for the first surely manifests the irrational belief that inanimate things are animate and can be punished, whereas the second has no such flaw. But behind this seemingly innocuous observation lies, as I shall argue, a false account of action explanation and a false semantic theory. According to the standard account of actions and their explanations, intentional actions are actions done because the agent has a certain desire/belief pair that explains the action by rationalizing it. Any explanation of intentional action in terms of an appetite or occurrent emotion (which might appear to be an explanation solely in terms of desire) is hence assumed to be elliptical, implicitly appealing to some appropriate belief.[1] In this paper, I challenge this assumption with respect to the "arational" actions of my title—a significant subset of the set of intentional actions explained by occurrent emotion. These actions threaten the standard account, not only by forming a recalcitrant set of counterexamples to it, but also, as we shall see, by undercutting the false semantic theory that holds that account in place.

I define these actions ostensively by means of a list of examples, and then define them explicitly, thereby making it obvious why I call them "arational" actions (rather than "irrational," on the model of

---

* Earlier versions of this paper have been read at philosophy colloquia at UCLA, UC/Irvine, and, most recently, at a conference on *Reason and Moral Judgment* at Santa Clara in 1989. I am grateful to the many people who contributed to the discussion on these various occasions, and also to Anne Jaap Jacobson, Christine Swanton, and Gary Watson for detailed comments on earlier drafts.

[1] So, for instance, Donald Davidson has said in lectures that 'She fled out of fear' (or 'because she was frightened') and 'She killed him out of hatred' are to be construed in terms of the actions' being caused by appropriate beliefs and desires.

0022-362X/91/8802/57–68

the distinction between "amoral" and "immoral"). I cluster the examples around the emotion (or emotions) that would, usually, explain the actions; the explanation would, usually, be of the form "I ϕ-ed because I was so frightened (or happy, excited, ashamed . . . so overwhelmed by hatred or affection or . . .) that I just wanted to, or felt I had to." Arational actions:

(a) explained by a wave of love, affection, or tenderness—kissing or lightly touching in passing, seizing and tossing up in the air, rumpling the hair of, or generally messing up the person or animal one loves; talking to her photograph as one passes, kissing it;
(b) explained by anger, hatred, and sometimes jealousy—violently destroying or damaging anything remotely connected with the person (or animal, or institution) one's emotion is directed toward, e.g., her picture, letters or presents from her, awards from her, books or poems about her; the chair she was wont to sit in, locks of her hair, recordings or "our" song, etc.;
(c) explained by anger with inanimate objects—doing things that might make sense if the things were animate, e.g., shouting at them, throwing an "uncooperative" tin opener on the ground or out of the window, kicking doors that refuse to shut and cars that refuse to start, tying towels that keep falling off a slippery towel rail on to it *very* tightly and then consolidating the knots with water; muttering vindictively 'I'll show *you*', or 'You *would* would you';
(d) explained by excitement—jumping up and down, running, shouting, pounding the table or one's knees, hugging oneself or other people, throwing things;
(e) explained by joy—running, jumping, leaping up reaching for leaves on trees, whistling or humming tunelessly, clapping one's hands;
(f) explained by grief—tearing one's hair or clothes, caressing, clutching, even rolling in, anything suitable associated with the person or thing that is the object of grief, e.g., pictures, clothes, presents from her (cf anger above). (The example of rolling in comes from a novel in which a man takes his dead wife's clothes out of the wardrobe, puts them on the bed and rolls in them, burying his face in them and rubbing them against his cheeks);
(g) explained by shame—covering one's face *in the dark*, or when one is alone; washing with violent attention to scrubbing and scouring;
(h) explained by horror—covering one's eyes when they are already shut;
(i) explained by fear—hiding one's face, burrowing under the bed clothes;
(j) explained by feeling proud, or self-satisfied, or pleased with oneself —talking to or posturing to oneself in the mirror.

I maintain, with respect to these examples, that on very many (though *not* necessarily all) occasions on which such actions were performed, it would be true to say the following of them: (i) that the action was intentional; (ii) that the agent did not do it for a reason in the sense that there is a true description of action of the form "*X* did it (in order) to . . ." or "*X* was trying to . . ." which will "reveal the favorable light in which the agent saw what he did,"[2] and hence involve, or imply, the ascription of a suitable *belief;* and (iii) that the agent would not have done the action if she had not been in the grip of whatever emotion it was, and the mere fact that she was in its grip explains the action as much as anything else does.

I shall say that when and only when these three conditions hold of an action it is, by definition, an *arational action,* and appropriate to the emotion or emotions that explain it. The examples are of action types, most of whose tokens would be arational actions, but, as noted above, I am not insisting that they always would be. Many of them, for instance, might be done on occasion without the agent's being aware of what she was doing, thus violating condition (i), and many might be done, on occasion, in order to . . . , thus violating condition (ii). On such occasions, the actions are not, as performed, arational actions; whether or not an agent has performed an arational action on some occasion is determined by whether the three conditions obtain.

I have encountered great resistance, both explicitly in discussion and implicitly in philosophy-of-action literature, to the very idea of arational actions as defined. What people want to do is deny that when condition (i) obtains, condition (ii) *can* obtain. "If an action is intentional," they say, "it *must* be done for a reason, i.e., because of an appropriate desire and belief"—for this is, indeed, the standard account. Now I do not want to quarrel about senses of 'done for a reason'; the central point at issue is certain belief ascriptions to agents performing intentional actions of the sort described above. I am just using 'not done for a reason' here to capture my claim that these actions are explained solely by reference to desire—"I was so angry/delighted, etc., I just wanted to"—not to an appropriate belief.

To get quite clear about what is at issue, let us consider as an example, Jane, who, in a wave of hatred for Joan, tears at Joan's photo with her nails, and gouges holes in the eyes. I can agree that Jane does this because, hating Joan, she wants to scratch her face, and gouge out her eyes; I can agree that she would not have torn at the photo if she had not believed that it was a photo of Joan; and if someone wants to say, "So those are the reasons for the action," I do not want to quarrel, for these "reasons" do *not* form the appropriate desire-belief pair assumed by the standard account. On the standard account, if the explanatory desire in this case is the desire to scratch Joan's face, then the appropriate belief has to be something absurd, such as the belief that the photo of Joan *is* Joan, or that scratching the photo will be causally efficacious in defacing its original. And my disagreement is with adherents of the standard account, who must think that some *non*absurd candidates for appropriate beliefs to ascribe to agents performing arational actions are available.

---

[2] "Explanations of action in terms of reasons work by revealing the favourable light in which the agent saw what he did (or at least what he attempted)"—John McDowell, "Reason and Action," *Philosophical Investigations,* v, 4 (October 1982): 301–5.

An exhaustively detailed rebuttal of the various candidates that may be offered cannot be given, of course, but I now review the most plausible.

When one attempts to think of appropriate beliefs, it becomes clear quite rapidly that there is no point in trying to find them piecemeal, token by token—beliefs such as the belief that by harming a photo one harms the original. Viewed abstractly, the desires to perform arational actions when in the grip of an emotion provide, apparently, a rich fund of those cases which Gary Watson[3] neatly characterized as those in which "one in no way values what one desires" (*ibid.*, p. 201), and what is needed to show this is mere appearance, is a belief, ascribable to the agent in every case, about the value of what she is doing. My claim was that these actions are not done in order to ϕ; the counterclaim, coming from the standard account, should be that they *are* done in order to ϕ, where the agent always wants to ϕ and where the appropriate belief, showing in what way the agent values the action, would be of the form "and the agent believes that doing this (gouging out the eyes on the photo or whatever it is) is ϕ-ing."

We seem to find the promise of a candidate in the thought that tokens of the types in question are done *in order to express the emotion,* to relieve it, or vent it, or make it known. This, it is said, reveals the standard desire-belief reason for which arational actions are done: the agent desires to express her emotion, and believes that whatever she is doing *is* expressing it.

But quite generally, what is wrong with this suggestion once again is that it involves ascribing a belief to the agent which should not be ascribed. If I ϕ in order to express or relieve my emotion, I do so in the belief that my ϕ-ing will indeed have (or is likely to have) this upshot. And in such cases, there is the possibility that I am not setting about fulfilling my intention in the right way; that I am open to correction. But arational actions would not usually admit of any possibility of mistake in this way; they are not the sort of action an agent would usually do in the (possibly erroneous) belief that they would achieve this effect. Nor should we accept it as obvious that in every case the agent has the desire *to express this emotion,* a desire whose content is distinct from that of the desire to, say, throw the tin opener violently on the floor. The ascription of this extra desire requires an extra justification.

---

[3] "Free Agency," this JOURNAL, LXXII, 8 (April 24, 1975): 205–220.

Of course, I grant that, on some particular occasion, such a justification may be available; as before, I am not insisting that *no* token of an arational action type is ever done in order to express an emotion. I might indeed, on occasion, break up the furniture in the belief that this is expressing my rage and wanting to express it, because my psychotherapist has convinced me that I have hitherto suppressed my emotions too much, that it is better to express them than to bottle them up. Here, my decision to follow the therapist's advice provides the justification for ascribing to me the extra desire *to express this emotion*—and, significantly, it also introduces the possibility of my being open to correction. "You're not really expressing your rage," knowledgeable onlookers may say. "Why don't you really *scream*!"

Similarly, I might tousle someone's hair in the belief that this will make it known to them that I am feeling a wave of affection for them, and that they would like to know this. Or I might deliberately try to get my corrosive hatred for Joan out of my system and tear at her photo in the belief that doing so will bring me this relief. In such cases, the information that my belief is false, that I am not going to succeed in making known or relieving my emotion, or that my expressing it in this way on this occasion is not a good idea would be seen by me as a reason for stopping what I am doing. Granted, this can happen on occasion. But on most occasions, this is not how things are; usually the agent will not be *trying* (successfully or unsuccessfully) to express her emotion, in the belief that this is a good idea, at all.

Notwithstanding this, there seems to be something right about the idea that arational actions are done "to express the emotion" that has not yet been brought out. Let us consider the claim, not that the agent does the action *in order to* express the emotion, but rather qua expression of emotion.

What does this mean? If it is saying that arational actions *are* "expressions of emotion" or "expressive of emotion," this is, I think, obviously true, but it adds nothing to the claim that arational actions are as defined—i.e., intentional actions appropriate to certain emotions, whose only explanation is that, in the grip of the relevant emotion, the agent just felt like doing them. Indeed, it is less explicit than the definition, for both unintentional actions (such as unknowingly gnashing one's teeth) and intentional actions done for further reasons (such as tearing at someone's eyes in order to hurt them) also count as expressions of emotions. But if the 'done qua expression' turn of phrase is intended to capture something more explicit, what could this be?

Given the obvious fact that in some sense the agent does arational actions "(just) because she wants to," or "for their own sake," prompted by the occurrent desire, it is natural to compare arational actions with actions prompted by appetite, and to look to Stephen Schiffer's[4] account of actions prompted by "reason-providing desires," such as eating a piece of chocolate because I am seized by a desire for chocolate. These, according to Schiffer, are done "for pleasure"—not in quite the same sense in which one goes to the ballet "for pleasure," where I desire to go to the ballet because I believe I shall enjoy it, but still in the sense that they are done *in the belief* that satisfying the desire will yield pleasure. This being so, the obvious cases of actions prompted by Schifferian reason-providing desires can be given a standard belief-desire reason explanation; the agent desires physical pleasure, and believes that acting in accordance with the currently aroused appetite will yield this, and so acts. And similarly, it might be said, in the case of arational actions; the agent desires pleasure, and the appropriate belief, ascribable in every case, is that acting in accordance with the occurrent desire to do whatever it is—tear up the photo, cover one's eyes, throw the tin opener out of the window—will yield this.

But the difficulty with extending Schiffer's account in this way is that the ascription of the relevant belief is plausible only with respect to the standard bodily appetites. It is indeed "almost always" true, as he says, that the bodily appetites, once aroused, are pleasurable to satisfy, which is why we talk of the pleasures of food and drink and sex, why we have the concept of *physical* pleasure, and why actions prompted by appetite can so often be described as done "for pleasure." The belief that we shall get physical pleasure from actions done to satisfy such appetites may plausibly be ascribed to any one of us. But the ascription is either implausible or vacuous in other cases.

Suppose I am seized by a sudden desire to lick something furry, and do so because of that desire. Do I do so *in the belief* that doing so will yield me pleasure? A philosopher could indeed, parodying Thomas Nagel,[5] introduce the notion of a motivated belief; the description of an agent as doing something "because she wants to do it" simply entails that the agent believed that doing whatever it is would "give her pleasure." This "motivated belief" could then be ascribed to the person who licks something furry "because she wants to"; but in such a case the ascription is clearly vacuous. There are no grounds on which she could believe it will give her *physical* pleasure —quite the contrary, in fact. She does not necessarily believe she will enjoy it—indeed, if one were seized by such an odd desire, one might well act on it because one was curious to find out whether doing so was enjoyable or not. The only "pleasure" the agent believes in is "the 'pleasure' of desire-satisfaction," and this is an entirely formal and empty concept of pleasure.

---

[4] "A Paradox of Desire," *American Philosophical Quarterly*, XIII, 3 (1976) 195–203

[5] *The Possibility of Altruism* (New York: Oxford, 1970), pp. 29–30.

Now, the desires to perform arational actions (unlike the aroused bodily appetites but like the desire to lick something furry) are not generally known as being pleasurable to satisfy; on the contrary, we know of some of the cases that acting in accordance with the desires makes one feel terrible, and of others that acting in accordance with them is neither pleasant nor unpleasant. To ascribe to the agent of an arational action, in every case, the belief that satisfying this desire now will yield pleasure is hence, implausibly, to ascribe an absurd belief—one for which the agent has no grounds, and which she knows is probably false. Or it is the vacuous ascription of a "motivated" belief.

Actions prompted by odd physical cravings are, I claim, genuine examples of cases in which "one in no way values what one desires" and are thereby counterexamples to the standard account of intentional action, albeit so odd and rare that they might be dismissed as fringe cases. Arational actions, however, are not, in the same way, odd or rare; if they do indeed resist appropriate belief ascription, as I have maintained, then the standard account is shown to be fundamentally flawed.[6]

For not only do arational actions provide a large set of counterexamples, but they also, once their resistance to belief ascription is acknowledged, justify our looking with a skeptical eye at actions done in the grip of an emotion to which the full rational panoply of belief, desire, and "intention with which" is usually ascribed. It is generally said, for example, that, if someone flees in terror from a lion, she is doing so in order to get to safety and preserve herself from danger or death; she desires self-preservation and believes that flight is the best way in the circumstances to get it. And, it is said, if someone strikes a person with whom she is angry, or says cruel things to him, she does so in order to hurt, or even to punish him; once again, an appropriate desire-belief pair is ascribed.

---

[6] I am not, of course, unaware of further variations on the appropriate belief which might be tried, but only space prevents me from showing that they fail, too. The belief that in performing the arational action one will eliminate discomfort or agitation (rather than achieve actual pleasure) does not turn the trick; nor does the fascinatingly symbolic nature of many of the examples of arational actions yield anything helpful, as many people are initially tempted to suppose.

Anyone who confidently holds the view that these ascriptions of reasons or "intention with which" must apply in such cases is committed to seeing a great disanalogy between them and cases such as feeling frightened of burglars, ghosts, or thunder and burrowing under the bed clothes (to safety?), feeling angry and kicking furniture (to hurt it?), and muttering imprecations under one's breath (for whose ears?), or to making them analogous by ascribing quite lunatic beliefs to the agent.

If there really is a great disanalogy, the account of the "rational" cases provides us with no clue about the account to be given of the arational ones, which must then seem utterly mysterious. Nor is the problem they present solved—not at least for anyone interested in giving a systematic account of action—by denying that arational actions are intentional. For they are clearly not unintentional, and to say they form a significant class of actions that are neither intentional nor unintentional is to admit that, within the standard account, they present a formidable problem.

If, on the other hand, they are accepted as analogous and the lunatic beliefs are ascribed, these will show up nowhere else in behavior, be sincerely and vigorously repudiated by the agent, and that agent's momentary acquisition of them will, in turn, be utterly mysterious.

A deep problem is found here by anyone who holds that whether or not an action is intentional or was done for a reason (because of a desire and a belief), whether or not an emotion was motivated by emotion *e*, and whether or not an agent believes that *p* must be all-or-nothing matters. And this view is held by people in the grip of the false semantic theory according to which predicates such as 'intentional', 'for a reason', 'motivated by emotion *e*', and 'believes that *p*' have clearly determinate, necessary and sufficient satisfaction conditions. On this theory, an action must be intentional or not, done for a reason or not, motivated by an occurrent emotion or not; an agent must believe that *p* or not. And then we find these mysteries.

The new "solution" to this problem is to say that, if the world proves thus recalcitrant to our attempts to carve it up with our predicates, this shows that there is something wrong with the predicates and the concepts they express; that, under the pressure of the facts about human behavior, they "fragment" or "come apart." So "intentional" and "belief" (for example) must be abandoned, and replaced by more accurate concepts derived from neurophysiology.[7]

But suppose we abandoned this false semantic theory, and instead said the following. Actions done because one is in the grip of an emotion do not form discrete groups, but a range. In the grip of an emotion, we do some things quite involuntarily, such as sweating, trembling, and coloring up. These are things over which, as things are, I have no direct control at all. There are other doings over which I can exercise direct control but can also do involuntarily, without realizing that I am doing them. I can clench or unclench my fists at will, smile or frown, but these are also things I can easily do unawares. I may begin to do some things, e.g., scream or cry or run without realizing that I am doing so, and once started, find it easier to go on than to stop; here perhaps I may be said to refrain intentionally from stopping. Other actions I do intentionally because, in the grip of the emotion, I just want to do them, though I do not do them in the belief that there is anything good about them at all (arational actions). Then there are actions that I do, momentarily believing that there is something good about them, though, looking back, I may not be able to understand how I could have (some cases of *akrasia*); and finally there are actions I do for a reason, in order to do or achieve something I believe to be good or desirable. Although there are clear cases in each of these groups, some of the actions that are clearly done "because the agent is in the grip of an emotion" will have features in common with some two adjacent groups, and there will be nothing that does or could settle to which group it "really" belongs.

Fleeing in terror might well be a case in point. It may be, on occasion, that one has a reason for fleeing—that it will get one out of danger—and flees, but it still may be that one did not flee, on that occasion, in order to get out of danger. Perhaps, after all, it is simply the case that one of the desires we are seized by when seized by fear is the desire to run or hide—such a desire would have good survival value—and that sometimes we act on it as impulsively, and with as little thought, as we act on the desire to scream or jump for joy.

Now, if this is indeed how things are with us, what are the consequences for the roles of reason and emotion in action? Well, let us go back to the thought with which I started—the contrast between the supposed irrationality of striking inanimate objects in anger and the supposed rationality of striking animate ones. If what I have argued is correct, then both sides of this contrast may be false. It may be that neither is irrational nor rational, but rather that both are arational, in the sense of being done without reason.

---

[7] Cf. Patricia and Paul Churchland's and Stephen Stich's writings, *passim*; e.g., Patricia Churchland, *Neurophilosophy: Toward a Unified Science of the Mind/Brain* (Cambridge: MIT, 1986), p. 382.

Moreover, insofar as one can see the potential for rationality or irrationality in either, the ascription goes the other way. I have deliberately stayed away from the murky waters of the topic of *akrasia* or weakness of will, but it is, I take it, perfectly obvious, and consistent with all I have said, that on some occasions an arational action may be irrational in the standard akratic way, i.e., contrary to some practical judgment about the good or necessity of refraining from it. If I throw the only tin opener out of the window, I certainly shall not be able to open the tin and may have to go hungry to bed; if I wreak violence on someone else's valuable antique furniture, I violate her rights. So, assaulting inanimate objects in anger may be irrational in the sense of being akratic. But it is surely the case that assaulting animate ones is much more likely to be so. Reason may well have nothing to say against assaulting the inanimate, but the fact that the animate can be hurt and harmed always stands as a potential reason against assaulting them.

In highlighting this point, I do not mean to ally myself with those who see emotion as opposed, in some important sense, to reason (or Reason) or who think that the practical rationality manifested in moral action must be somehow independent of the desiderative faculty. On the contrary, I stand firmly on the Aristotelian side of the Aristotelian-Kantian debate. But those of us who follow Aristotle should not, I think, push our luck too far, and I want to conclude by showing how the existence of arational actions creates something in the way of a problem even for us.

An important fact about human beings, stressed in neo-Aristotelian virtue-based ethics, is that we are creatures such that our appetites and passions may prompt us where reason would also have lead. In this fact lies the sense in which we are "constituted by nature to receive the virtues,"[8] the possibility of harmony between our desiderative and rational faculties, of the virtuous person's grasp of "truth in agreement with right desire" (*ibid.*, 1139 a31). Aristotle maintains, (and Dennis Stampe[9] has recently reiterated) that desire is for the seeming (or apparent) good (or pleasant); when what seems good in the faculty of desire is truly good, then the desire is right, and the true judgment of reason about what is good will be in agreement with it.

---

[8] *Nicomachean Ethics*, 1103 a24–6.

[9] "The Authority of Desire," *Philosophical Review*, XCVI, 3 (1987). 335–81.

Now, the apparent pleasures to which the bodily appetites prompt us may indeed be judged truly good and pleasant by reason. In relation to the bodily appetites, Watson rightly emphasizes the fact that we may "judge that to cease to have such appetites is to lose something of worth" (*op. cit.*, p. 213) and thereby both desire *and* value (some of) the actions to which the appetites prompt us. A human being *can* be seen "in his role as Rational Animal"[10] while eating, drinking, and making love (if it is in the right way, on the right occasions, and so on), odd as that may initially seem, not because he is, necessarily, acting "for a reason" of the standard sort, but because these are activities that can properly form part of a flourishing human life; reason may correctly judge that such actions are good and endorse them.

The same is true of most of the emotions. Reason may judge truly that to cease to have many of the emotions to which we are subject would be to lose something of worth, thereby conferring value on many of the actions to which they prompt us which it would endorse. Once again, a human being can be seen in her role as a rational animal when she flees the dangerous, honors the dead, repels aggressors, punishes wrongdoers, makes recompense for her own wrongdoing, cherishes her children, celebrates joyful occasions, and so on. Reason may correctly judge that such actions are good.

The only sense in which Reason and emotion are opposed, according to Aristotelian ethics, is that (except, perhaps, in people who have "natural virtue") the untrained passions tend to represent things as good and pleasant (or bad and unpleasant) which are not truly so. We have to be trained to fear dishonor more than death, to desire sexual intercourse as an expression of love within a lifelong partnership rather than as simply fun or the exercise of power, to be angered by injustice rather than deserved criticism. Unless they are properly trained, the passions will prompt us to action contrary to reason, or, even worse, corrupt our reason so that we judge things to be good and pleasant falsely. But, properly trained, most of them—it has been said, all[11]—will be in agreement with rational judgment.

---

[10] Davidson, "Actions, Reasons, and Causes," this JOURNAL, LX, 23 (November 7, 1963): 685–700.

[11] In "Aristotle's Doctrine of the Mean," *American Philosophical Quarterly*, X (1973): 223–30, J. O. Urmson maintained that, according to Aristotle, there is no emotion that one should never exhibit. In "Plato on the Emotions," *Proceedings of the Aristotelian Society*, Supplement, LVII, 1 (1984): 81–96, I argue, following Aquinas (*Sum. Theo.* 1a2ae Q24 a.4.), that, on the contrary, some passions may be bad in themselves, insofar as they involve an attachment to the truly bad or an aversion to the truly good, for example, envy and *accidie*. But the intrinsicially bad emotions are few.

But can the same be said about the arational actions to which our emotions prompt us? It seems to me that, by and large, it cannot, though we might make out a plausible case for some of the actions appropriate to a wave of affection or tenderness. Reason can find good in touching and caressing the people and animals one loves (in the right way on the right occasions, etc.); it forms and endorses loving bonds, is found innocently pleasurable, reassuring, or endearing by the recipients, may speak louder than words, especially to the pre- or nonlinguistic, and so on. Although we usually do not rumple our children's hair for these excellent reasons, we could. But I do not see how one could even begin to make out a case for finding any good in any of the others; they are arational not merely in the sense that one is prompted to them only by desire and not by reason (which is equally true of actions prompted solely by appetite) but further, in the sense that they cannot be *made* rational; reason cannot endorse them.

We might still think, however, that our lives would not, quite generally, be better if emotion never prompted us to act in these ways, or if we always resisted the prompting. We might well find something rather touching or endearing about people's performing many of the arational actions;[12] even the disturbingly violent ones seem to evoke some sort of bond of sympathy. When I have read this paper to discussion groups, I have found that the list of the examples at the beginning always provokes instant delighted recognition; everyone knows what it is like to act in some of these ways, and is somehow pleased to hear it acknowledged and described. Now, someone might maintain that this is just a case of the weak and fallible taking (improper) pleasure in having company. But to me it suggests that we value ourselves and each other as emotional creatures—not as *rational*-emotional in the way pinpointed by Aristotle, but as just plain emotional—and do not believe that the perfect human being would never act arationally.

The importance of this fact, if it is a fact, should not be overemphasized. It leaves general claims about the connections between human perfection, moral agency, and practical *rationality* intact, *if* it is remembered that we make these claims in the area where things are, as Aristotle says, not true of necessity but "for the most part."

ROSALIND HURSTHOUSE

The Open University

---

[12] I owe this point to Watson in discussion.

SUBSTANCE: PROLEGOMENA TO A REALIST THEORY OF IDENTITY

DIFFERENT philosophers have used the word 'substance' in different ways, and it may therefore seem that there is no one thing with which they were all concerned and which we can now try to characterize correctly. Yet the admittedly rather various employment of the term has in general constituted, among other things, a sustained attempt to understand (if also sometimes to improve) fundamental ways of thinking reflected in the deep structure of all natural language. There have been several disparate sources of hostility to the notion of substance. The old-fashioned empiricist's horror of supposed unknowables is perhaps a thing of the past, replaced by the more fashionable neo-idealist's or conceptualist's distaste for explanations of specific logical form which present it as other than arbitrary, mutable, and at best useful. An important motive is the feeling that modern physics has undermined the claims to ultimacy inherent in the substance-based ontologies of the past, as any currently preferred ontology is in turn vulnerable to future physics. The aim of the present paper is to show how traditional realist doctrines, suitably interpreted, have an explanatory force that pragmatism or any other form of conceptualism cannot hope to match. The explanations it supplies are of structural features of our thinking to which considerations drawn from the philosophy of physics are simply irrelevant.

Although preconceptualist realists disagreed over what things are paradigm substances, as well as over the details of what it is to be a substance, it is not difficult to extract from the tradition at least a rough list of the category's leading characteristics. Such a list of properties should even help us to map and understand the disagreements and divergences. Yet orthodoxy and broad principles are more relevant to the immediate purpose than heresy and idiosyncratic variations. The present argument will therefore assume that such things as horses and plane trees are paradigm substances, whereas homoeomerous substances such as gold and water call for special, but integrally related treatment (treatment sketched out below under the heading "materiality"). Attributes, such as a thing's redness or squareness, and events, actions, or processes, such as walks, walking, thunderstorms, and the like, exemplify nonsubstances or (for the sake of a single name) "modes."

Most of the claims embodied in the following list have on occasion been impatiently brushed aside, or simply ignored, by modern phi-

0022-362X/91/8802/69–90

losophers arguing on behalf either of empiricism or of conceptualism. Some of them, however, have been reinterpreted and absorbed into this or that version of conceptualism. One aim of the present argument is to reveal the coherence of the list, which will emerge in so far as each item can be explained, and needs to be explained, by reference to the others. Some effort will be made, of course, to present the principles as tenable, although more than one principle will be modified after consideration of the motives for which it has been held.

1. Substances are the ultimate subjects of predication, and therefore the only beings with independent existence.
2. Substances are real unities (both natural and logical).
3. Substances are material. Individual substances are distinguished from one another at any one time by their matter.
4. Substances exist all at once, and exist through time, or endure. (Events, in contrast, take time or unfold.)
5. Substances are active, the ultimate sources of change. Their underlying natures or essences are the ultimate principles of explanation.
6. Only substances fall into true natural kinds, and every truly individual substance is a member of a natural kind.

Let us consider these candidate properties of the category *seriatim*.

I. LOGICAL PRIORITY AND INDEPENDENCE

That substances are the ultimate subjects of predication is perhaps the most famous claim made about the category. The other categories are predicable of substance, whereas substance is only predicable of itself. There are, of course, particular subjects of predication not in the category of substance: the color of this curtain, for example, which is the subject of the assertion that this brown is fading fast (universals do not fade). Yet such subjects owe their existence to the satisfaction of a predicate by something else. This brown exists just as long as the curtain is brown. My run goes on as long as I am running. It is admittedly also true that this horse exists just as long as something is a horse: yet here the 'something' is not something else, but is the horse itself. Their having no need for something else as their subject is what has given rise to the ontological doctrine that substances are the only logically independent individuals possessing independent existence.

A nonsubstance or mode may exist dependently on another nonsubstance, as beauty may belong to a shape. But then the shape, as a particular existent, will normally depend on something shaped, and ultimately every mode depends on a substance or substances. A shape, it is true, may sometimes exist as a particular without belonging to a particular thing or substance. The triangle formed by three pencils on a desk exists, although no substance on the desk is triangular. But then it exists just as long as the pencils are in a certain relation.

Logical and metaphysical dependence is taken to be reflected, if imperfectly, by the grammatical phenomenon of nominalization. Broadly speaking, general nouns that are not substance nouns are nominalizations of predicates that are not nouns. The imperfection of this rule of thumb is easy to illustrate: 'courageous' derives from 'courage'. Other counterexamples seem more interesting, because less accidental. Although the noun 'square' may be presumed to derive from the adjective (and so, too, with the French '*rond*'), 'triangular' is a grammatical offshoot from 'triangle', 'circular' from 'circle' (or '*circulus*'). Yet, even if grammar and etymology reflect logic imperfectly, the logical point is itself unaffected. Whether or not the verb 'battle' is linguistically precedent to the noun, a particular battle exists just in so far as certain people are fighting.

The question between realism and conceptualism is whether that logical point reveals something about the different ontological status of people and battles, or else merely reflects an accidental feature of "our conceptual scheme." It is true that the advocate of conceptual relativism is likely to reject this characterization of the issue on the grounds that ontology itself is always relative to a conceptual scheme: ontological assertions, it is commonly held, express preference for one scheme over others or (if talk of alternative schemes is thought to be beyond the bounds of sense) purport to reveal something about the only scheme we have. But the principles of pre-Kantian theories of substance, including the notions of dependent and independent existence, can only be understood in terms of a conception of ontology as absolute. For the conceptualist, all beings are concept-relative: for the realist, some beings are concept-relative but others, substances, are not.

Acceptance of the famous logical criterion of substantiality, when interpreted from a realist point of view, is simply equivalent to the doctrine that an individual substance is a natural or real individual, in some sense "given" and not "constructed." The term 'horse' simply *classifies*, or gives a general name to, natural individuals, the horses themselves. The adjective 'brown', on the other hand, *characterizes*, not browns themselves, but brown objects. Particular browns are sliced out of reality by the concept *brown* or, if you like, by the nominalized predicate, 'brown' or 'brownness'. Such entities are *abstract* particulars, mere aspects of things. Their individuality is posterior to their individuation by us, or by our concepts. But the individuality of horses and the like is prior to their being picked out and classified. As Aristotle says, both separability and individuality belong primarily to substance. Hence the thought that runs so deeply through the preconceptualist tradition, that nonsubstances or modes, in being logically and ontologically dependent on substances, are at the same time in some sense mind-dependent or mind-created. Whereas individual substances are individuated by a real distinction, modes are individuated by a distinction of reason. It is that kind of mind dependence which modern conceptualism wishes to make universal, that kind of distinction between distinctions which, in effect, it sets out to face down. In this it has been remarkably successful, if success is to be reckoned solely by the number or eminence of its adherents.

II UNITY

The principle that substances are real or natural unities identifies what is perhaps the most fundamental feature of the category of substance, as well as one of the most controversial and many-sided. The notion of substantial unity has itself seemed alarmingly liable to fragmentation, and shifts in its interpretation are central to the story of early-modern metaphysics.

The topic of unity can be approached, first, through a view of the substance as the one simple subject of a multiplicity of predicates or accidents; and, second, through a view of individual substances as unitary and discrete physical wholes, each naturally and in itself distinct in a way in which abstract, nonsubstantial individuals are not. Admittedly, the question arises whether these approaches can be presented as approaches to the same topic only by an equivocation of the term 'unity'. How, after all, can a philosophical problem about (to put it crudely) what ties a thing and its various attributes together be one side of the same issue, the other of which is a matter of comparing different sorts of principles of distinction between individuals in different categories? Is the first problem, for example, not a problem that, if it arises at all, arises equally and in the same way with respect to all subjects of predication of whatever category? What connection does it have with the question of the natural unity of horses by comparison with processions? Yet, despite the prima facie force of these questions, it may well be that we can expect to understand predication only through understanding the most primitive or fundamental case of it, namely, that case in which sensible qualities are predicated of a natural or "given" subject. Such subjects, it may plausibly be argued, are there to be picked out by sensory means prior to predication, and are to be contrasted in this respect with subjects that are picked out only through nominalized first-order or higher-order predicates. So the topic of the subject-predicate relation and the topic of individuation may, after all, be rather closely connected.

Many things could be said about unity, but I want only to draw attention to two traditional points of contrast between substances and nonsubstances which seem to identify extremely interesting and significant logical phenomena crying out for explanation that modern theory simply does not give. The first relates to the absence of natural wholes and parts among nonsubstances as compared with substances, i.e., to the kind of *physical* unity of which substances alone are capable. The second relates to a certain lack of *logical* "wholeness" or unity from which nonsubstances suffer by comparison with substances.

(i) Parts and wholes. An attack of chicken pox may seem substance-like not only in possessing something like a Lockean (or Kripkean) real essence, but also in having a history for the most part internally determined by its earlier stages. Yet there is no denying that its individuality is imposed by us, its boundary conditions or existence conditions are set by our concept. To see that this is so, consider a variety of possible candidates for the status of being a case of chicken pox. In different subjects, let us suppose, the virus is present as follows:

(a) without symptoms, beyond positive results to certain tests;
(b) with minor characteristic symptoms (such as swollen glands and a slight "cold") but without "illness";
(c) with fever, and so "illness," but with no other characteristic symptoms;
(d) with a spot or two, but no fever;
(e) with fever, spots and other characteristic symptoms (the paradigm case).

In which of these cases does the subject have chicken pox? Perhaps we exclude (a) and (b), since they are not cases of illness. We let in (d), however, since the spots are so characteristic. That does not tie 'having chicken pox' to symptoms, without regard to a presumed common aetiology. It is just that the "real essence" here, a certain virus's acting to a certain effect, owes its boundary conditions at least to the general, functional notion of being ill, if not to some more specific determination of the effects.

Here we may bear in mind how pragmatically the genus of *diseases* is determined. Yet this general notion essentially helps to determine, not only the boundary conditions of the species, but also the boundary conditions of the individual. In this respect, the genus *disease* is quite different from the genus *animal,* for the latter is part of a hierarchical system logically (even if not, in some minor respects, genetically) built up from below, *from* the level of given particulars. A mark of this difference is that problems about the boundaries of the genera *animal, mammal,* and so forth are in general quite distinct from problems about the boundaries of the species. If there is a group of individuals that, after exhaustive exploration and appraisal of their characteristics, continue to give rise to doubt as to their membership of the species *mountain gorilla,* their problematic status will not be due to doubt as to whether they are animals, or even whether they are primates. In cases where the question is purely one of taxonomy (i.e., where the doubt is not due to factual ignorance of, e.g., the physiology of the doubtful specimen), anything that is possibly a mountain gorilla will certainly be a primate, except in so far as it may be questionable whether mountain gorillas are best classified as primates. But it may be questionable whether someone has chicken pox just because he is not clearly classifiable as ill. It may be beyond question that, *if* he counts as being ill, then the attack of illness is an attack of chicken pox.

The imposed or conceptual character of the boundary of an attack of the disease shows itself again if we consider, for example, the place of the individual episode in the history of the victim, its relation to other aspects of the victim's general condition, and so forth. If the virus responsible for some disease is commonly present in the body before and after an attack, and is liable to cause a recurrence, then the attack appears as a merely striking slice, which is bounded only in so far as it is striking in a way that matters to us, out of a continuing process in the victim. If the illness occurred in the context of a general breakdown of the victim's health, it can equally appear as a part or aspect of that other process. If the victim infects others, the process taken to encompass what goes on in him may have the wider dimensions of an epidemic, or of the history of a strain of virus. Every episode is a part of indefinitely many wider episodes, and there are no natural wholes or parts, no "given" articulation of episodes. It is our concept that determines that the smoke is an effect of the fire, whereas the flame is a part of it. The only reason why the breeze now blowing over England is a result, rather than an element of the hurricane out to sea may be that it is too gentle to count for us as the latter.

By contrast, a horse is by nature a whole, its undetached hoof is by nature a part of that whole. To say that a horse is a natural whole is to say that it is not a part of anything in the way in which its natural parts are parts of it. A team of horses is not an individual substance, but a collection of individual substances. Collections have members rather than true parts: every part of a part of an individual is a part of that individual, but a part of a member of a collection is not a member of that collection. Only in the category of substance are there natural or "given" wholes or parts. We can regard a thunderstorm or a procession as a whole with parts, but every such event or process is a part of some wider event or process no less (if no more) unitary or distinct.

This thesis may be further tested against rather more substance-like nonsubstances than disease episodes, e.g., such entities as waves, fires, and hurricanes. Each involves a sort of principle of activity which, once on its way, seems relatively unaffected by the normal run of surrounding circumstances, and through which what occurs later is largely determined by what occurs earlier within the individual. A bit like a living creature, each has a normal, "internally" determined course. As David Wiggins says of substances in his book *Sameness and Substance,*[1] such entities are not in equilibrium with their circumstances. Yet the analogy with substances is only partial, and in the case of such modes there is a conceptual choice. We can think of a *fire,* as perhaps the Fire of London is normally thought of, as an occurrence with earlier and later stages which took place and took time. But we can equally take the process of combustion itself to be the fire (the word 'process', of course, shares in the ambivalence). So understood, a fire may spread from one place to another, may be fanned fiercer by the wind, and may be wholly or partly extinguished by a bucket of water. By a third way in which we can conceive of a fire, it may seem to be a true individual substance, a space-occupying, movable, enduring object, which exists all at once and is composed of the matter being consumed and continually replenished. That is perhaps the way Heraclitus thought of a fire. The same three sorts of possibilities arise in respect of *hurricanes.* A hurricane may be an event with a beginning, a middle, and an end. Otherwise, it may be a disturbance or activity with a spatial center, which moves about, increases or decreases in extent and ferocity, is "born," "lives," and "dies." Finally, we can conceive of a hurricane as something even more substantial, possessing a "body" of air and cloud, with changing, replaceable material parts. There are perhaps just two ways of thinking of a *wave:* either as a disturbance in a medium or as something quasi-substantial, more or less high, and composed of rapidly ever-changing water.

---

[1] New York: Blackwell, 1980.

These considerations could well provoke an argument hostile to the thesis that the individuation of substances, by contrast with the individuation of modes, is not at all arbitrary. For it might be argued that how we think of a hurricane or a fire is, after all, a matter for arbitrary decision. It might be claimed that, for each of these ambiguous predicates, there are certain satisfaction conditions common to its different senses; and that whether we think of the entity in question as an occurrence taking time, or as a mobile disturbance or activity with a life history, or as a material thing with a body of air or gases (or whatever) united by a principle of activity, simply depends on which "criterion of identity" or logic we decide to couple with the common "criterion of application." In so far as the decision seems arbitrary, it might appear that in general the "primacy" of the category of substance for us is nothing but an ingrained habit: it is customary and for some reason convenient to associate one particular type of criterion of identity with a broad class of nouns like 'horse' or 'rabbit' which mark important principles of activity. It simply needs an effort of imagination to see that, by analogy with 'hurricanes', we might just as well think of 'horses' as occurrences with temporal parts, or as processes. Yet this argument for "ontological relativity" does not explain why hurricanes and fires are rather poor individual substances as compared with horses, and the whole case

A	*B*
1. The hurricane (1), conceived of as an occurrence with temporal parts, etc.	The physical, internal, life history of the horse.
2. The hurricane (2), i.e., the mobile disturbance conceived of as existing all at once and subject to change and to different possible futures, etc.	The life process of the horse.
3. The hurricane (3), conceived of as a quasi-substance composed of ever-changing material (violently moving air and cloud) united and distinguished by a principle of activity (a certain sort of swirling motion).	The horse.

loses a good deal of its seeming force if we devote more than the briefest reflection to the supposed analogy, provisionally set out herewith.

To the conceptualist, it will appear an accident that such a name as 'horse' is employed for $B3$ rather than $B1$ or $B2$. The anticonceptualist intuition, on the other hand, and the realist's dissatisfaction with fires and hurricanes as individual substances might find expression in the claim that there is an evident disparity between columns A and B. The individuality of $B3$, the horse, is evidently prior to the individuality of $B1$ and $B2$. The horse is the "given" individual, whereas its life and life history are parasitically identified. In column A, no individual is "given," least of all $A3$. An event or process has no natural boundaries, no natural beginning or end, unless parasitically. Some water's coming to the boil, or freezing, is just a part or aspect, sliced out by us, of the continuous physical process involved in the water's possessing variable temperature. If the nature or essence of boiling or freezing or exploding strikes us as individuating a discrete event because its coming into being involves a change of kind, not merely of degree, then that is just the way it strikes us. But a horse really is a distinct individual. It strikes us as distinct and unitary because we apprehend it as it is.

These claims are opposed to those of Gottlob Frege,[2] whose pronouncements on individuation have prophetic status for some analytic philosophers. In expounding his blanket conceptualism with respect to the natural world, Frege appealed to the fact that "I can say with equal truth both . . . 'Here are four companies' and 'Here are 5,000 men'," but he evidently failed to reflect on the different ontological status of collections and individual substances. Accordingly, he regarded natural unity as in the end a psychological matter:

> The more the internal contrasts within a thing fade into insignificance by comparison with the contrasts between it and its environment, . . . the more natural it becomes for us to regard it as a distinct object. For a thing to be "united" means that it has a property which causes us, when we think of it, to sever it from its environment and consider it on its own (*ibid.*, p. 42).

For Frege, all unity is concept-relative: nothing in the physical world is individual or a whole, but thinking makes it so.

The notion of substances as unitary is evidently related rather closely to the principle that substances are the ultimate, logically independent subjects of predication. If individual substances are "given" individuals as individual modes are not, that is just *because* substances enjoy a kind of natural unity that modes cannot in principle enjoy.

---

[2] *The Foundations of Arithmetic*, J. L. Austin, trans. (New York: Blackwell, 1968).

(ii) Composite terms. The second point I want to make about unity, although closely related to the first, bears more directly on the subject-predicate relation. It is at least implicit in Aristotle's reflections about substance that the only real or natural ultimate species are species of substances, a principle which Locke[3] took brief note of with respect to modes:

> . . . in *mixed Modes* any of the *Ideas* that make the Composition of the complex one, being left out, or changed, it is allowed to be another thing, i.e., to be of another species, as is plain in *Chance-medly, Manslaughter, Murther, Parricide,* etc. (*ibid.*, pp. 500f)

In other words, if we define a new term by adding arbitrarily to the notion of a species of mode, we create a new species; but not if our definition adds to the notion of a species of substance. If we suppose that 'thunderstorm' were introduced as a new term, definable as *storm with thunder,* then thunderstorms would have no less right to be supposed to constitute a species than storms. But the concept *thunderstorm* could as easily be divided in turn. Thus, there are no "last species" of modes. Yet, if a "baker" is defined as *human being (homo) who bakes bread* or an "adult" as *fully-grown animal,* bakers or adults do not constitute a species, but are what Aristotle[4] called *syntheta,* combinations. The parts of the concept *baker* or *adult* do not, as it were, amalgamate: *baker* remains irremediably a composite concept incorporating, but not uniting, species plus accident.

Approached from this direction, the contrast needs to be drawn with some care and qualification, not least because of the doubtful objectivity or reality of the species and, still more, the genera of substances. Locke, indeed, was attacking the point in the light of his skepticism about any natural species at all. But the Aristotelian tradition fortunately supplies the material for another, rather less cluttered approach, through the topic of individuation.

An adult is a fully-grown animal, but the human adult is identical with the boy or girl who grew up, i.e., the adult began to exist when the human being began to exist, not when the human being became fully-grown. A baker is, by definition, a human being who bakes bread for a living, but that addition to *human being* has no effect with respect to the identity conditions or existence conditions of a baker. A baker who retires or changes occupation does not thereby cease to exist. By way of contrast, although 'downpour' may be defined as *heavy rainfall,* in particular cases the downpour and the rainfall may well be distinct, since the latter may have been taking place for hours before the downpour began (i.e., before the rainfall became heavy) and may continue to do so after the downpour is over. Similarly, the place of the rainfall may be larger than the place of the downpour. As we might put it, although 'downpour' is defined as *heavy rainfall,* the downpour is not the rainfall but the heavy rainfall: the parts of the definition amalgamate. Hence, it is only if the rainfall is always and everywhere heavy that it is identical with the downpour. Otherwise, the downpour is a part, temporal or spatial or spatiotemporal, of the rainfall.

---

[3] *Essay concerning Human Understanding,* P. Nidditch, ed. (New York: Oxford, 1975). In fact, actions raise special issues

[4] Cf *Metaphysics,* VII. 4.

Recent writers on identity are not unaware of the class of composite terms. One of the more perceptive, Wiggins, seems first to have noticed them in the case of what he called "phase-sortals" in his book, *Identity and Spatio-temporal Continuity.*[5] Since then he has made further use of the notion there introduced of concepts that are "restrictions" of underlying more general sortal concepts. In these, he says, "the principle of individuation is certainly present, but it is buried and must be disinterred" from what is irrelevant to it. The principle of individuation associated with (his examples) the terms 'boy' and 'Anzac' is supplied by the "buried" concept *man.*[6] Whether or not the notion of a restricted sortal offers an adequate account of composite terms (and it seems clear that it does not), it should be recognized that the bare existence of Wiggins's examples present a twofold problem of fundamental importance to the theory of identity. First, *why do clear instances of such terms exist only among noun predicates predicable of substances* (or perhaps, to leave the more general possibility open for the present, of substance-like continuants)? And second, *what constraint here prevents us from manufacturing individuals by arbitrarily tinkering with principles of individuation*? In other words, what prevents us from introducing a substance term, say 'bakist', to express the concept *human being who bakes,* while yet intending this definition in such a way that the sense of the term *does* contain a principle of individuation distinct from the principle associated with the term 'human being'? What prevents the parts of the concept from amalgamating or uniting in the same way as the parts of the concept *downpour*?

---

[5] New York: Blackwell, 1967.

[6] *Sameness and Substance,* pp. 63–6; cf. *Identity and Spatio-temporal Continuity,* pp. 30–4, 59–63.

Now, these questions must either be shown to be based on false presuppositions, or it must be accepted that it is a shortcoming of current theory of identity that absolutely nothing is done to answer them. Any proposal that the concept *baker* is individuative only in virtue of its including or restricting or entombing the concept *human being,* which is individuative per se, whereas the concepts *rainfall* and *downpour* are both individuative per se, is of course no answer. It is merely a redescription of the problem in terms favorable to conceptualism, a pseudo explanation consonant with the assumption that special "concepts" with individuative power are responsible for slicing out individuals.

The best bet for the conceptualist might therefore seem to be to reject the presuppositions of our two questions. But can they be rejected? Well, with respect to the first question it might be argued that it is only through a kind of linguistic accident that there seem to be no clear analogues among nouns predicable of modes for 'girl', 'bachelor', and 'baker'. Suppose it to be stipulated that 'riot' means *disorderly assembly*. Then the time at which the riot began (it might be conceded) would correctly be taken by English speakers to be the time at which the assembly became disorderly. Yet why (it might be asked) should we not understand the definition in such a way that the riot *is* the assembly, and began when the assembly began even if the assembly began and long continued as a peaceful demonstration? The rough answer seems to be that it is impossible to make sense of the suggestion that a riot was going on when nothing satisfied the predicate 'riot', i.e., nobody was rioting. By contrast, the parallel claims that the baker existed before anything satisfied the predicate 'baker', and that the caterpillar continues to exist after it has ceased to satisfy 'caterpillar', are logically impeccable. Note that nothing is gained by stipulatively defining a term, e.g., 'orgy', in some such way as *party at some time in the course of which such and such a degree of impropriety occurs*. That definition might achieve by force of stipulation an unnatural answer to the question 'When did the orgy begin?', but it would not achieve an analogue to 'baker'. For under it, a party that begins with the guests' respectably sipping sherry but ends with large-scale impropriety over the sauternes would satisfy the predicate 'orgy' *throughout* its existence, whereas a man satisfies the predicate 'baker' only while his occupation is the appropriate one.

In order to reject the presupposition of our second question, on the other hand, it would be necessary to deny, in the face of the logical phenomena, that there are any constraints on the manufacture of principles or criteria of identity for substances. That is the response to be expected from the mainstream conceptualist, for it is an unargued and unquestioned assumption of many writers on identity that we can arbitrarily stipulate modifications of the "criteria of identity" associated with substance terms, just as (they suppose) we can stipulate "criteria of application": and that if anything restrains us, it is no more than ingrained conceptual conservatism. But if there is such a thing as an obviously false claim in philosophy, then that is such a claim. Nevertheless, let us just consider what happens if we stipulatively insist that a "bakist" is a being that exists just as long as a human being is a baker. What kind of thing is a "bakist"? With those existence conditions it inevitably turns out not to be substance, or a material thing of flesh and blood, at all. The "bakist" that exists just as long as Jones is a baker has to be such an entity as Jones's career as a baker, or his practice of bakery. If we stipulate existence conditions, the resultant entity has to be such that its existence conditions *can* be stipulated, i.e., it has to be a mode, it has to change categories. A downpour, on the other hand, is categorially no different from a rainfall.

It remains to consider such connections with other attributes of the category of substance as may help us to understand the phenomenon of composite terms. It might be argued that their possibility is simply an aspect of the *duration* of substances, i.e., of substances' lacking temporal parts and existing all at once. For it is only in the context of the false doctrine that a substantial object has, or may be regarded as having, temporal parts that it can be at all plausible that such predicates as 'baker' or 'boy' pick out individuals distinct from the human being. For these supposed individuals, it can then appear, share matter with the human being as unproblematically as his hands and fingers share matter with the human being. Moreover, there are nonsubstantial "continuants," such as nations and whirlpools, which have borrowed the logic of substances to the extent that they exist all at once and may thereby seem as capable as substances of sustaining composite predicates. The possibility that a nation should be for a time a debtor or at war does not prompt us to distinguish the debtor or the combatant from the nation.

Nevertheless, there are reasons for rejecting the conclusion that a full explanation of composite terms is to be found in the duration of substances or, more generally, of "continuants." First, there is the point that the spatial, as well as temporal extent of individuals is in question. Second, in the case even of the most substance-like nonsubstantial continuants, the logical constituents of their quasi-composite predicates may tend to amalgamate as those of composite predicates never do. A kingdom is a nation or state ruled by a king, but we may nevertheless suppose that the Kingdom of Italy ceased to exist, as the Italian state did not, with the formation of the Republic. It is true that we may prefer to identify the kingdom with the republic, and both with the state. But there is no requirement to do so, as there is to identify the baker with the man or, if he changes his job, with the butcher. The continuity of the state is no more "given," or less notional, than the extinction of the kingdom. To a nationalist, the nation may seem the more natural individual, to a monarchist, the kingdom. But that issue is political, not ontological. So too with nonsubstantial continuants less dependent for their existence on human convention. Suppose that 'hailstorm' is stipulatively defined as *storm with hail.* Suppose, too, that both storms and hailstorms are conceived of, not as events, but (with hurricanes and tornados) as meteorological disturbances that move from place to place and undergo change, i.e., as substance-like continuants. Nevertheless, there is no constraint on allowing that a particular hailstorm has had a shorter life and narrower extent than the more general storm of which for a time it has formed a part.

From such examples as these, it seems clear that the amalgamation of the parts of a definition reflects the notional status of the individuals denoted by the terms defined, whereas the logical phenomenon of composite terms or Aristotelian *syntheta* derives from the status of substances as natural or given individuals. For nonsubstances or modes, as we have seen, classification is prior to individuation and (to be rough) is relative to a predicate predicable of something else. Particular modes are abstractions sliced out of reality by general concepts. It is therefore unsurprising that modification of the predicate is liable to breed fresh individuals. Particular substances, however, are naturally discrete individuals whose given individuality is prior to their being identified and classified by us. It is therefore unsurprising that mere addition to the classificatory predicate or general name we employ will *fail* to breed fresh individuals. That explanation, in contrast to empty conceptualist redescription, looks like a real explanation of an ineluctable logical constraint.

III. MATERIALITY

The doctrine that substances are compounds of form and matter was not based on nothing. Paradigm substances drawn from the ordinary objects of experience are indeed organized matter. Exemplary immaterial substances (the Prime Mover, the active intellect, the Christian God, immortal souls) are all metaphysically queer, beset with difficulties in their conception. It will therefore be assumed that notions of immaterial substance are aberrant and incoherent, and that all substances are material objects.

The materiality of substances is evinced in the division of the category of substance into "things," such as dogs or oak trees, and homoeomerous substances or stuffs, such as gold or water. This is a division of types of substance corresponding to which there are two types of individual: on the one hand, material objects and, on the other, such quantities of stuff as the gold in this ring or all the tea in China. The latter are perhaps dubious "individuals," but, however that may be, they are certainly not "given" as such, since the boundary of whatever quantity of stuff we wish to identify is up to us.

Modern conceptualist doctrine is ambivalent in its attitude toward materiality. On the one hand, most conceptualists are prepared to talk without qualms of a category of "material objects" and even to allow it logical primacy within the conceptual scheme we have (although for some the honor is shared with events, or even with numbers). On the other hand, formal explanations of the thing/stuff dichotomy are commonly advanced at a level of abstraction that carries it up out of the category of substance altogether. So-called "count nouns," among which are 'walk' and 'laugh' as well as 'horse', are opposed to "mass terms," which include 'walking' and 'laughter' as well as 'snow'. If so highly abstract an approach were successful, i.e., if it could give us an explanatory account of the thing/stuff distinction, then such entities as laughs and battles would have every right to be regarded as material objects. Battles would be composed of fighting, or of such more specific stuffs as bombarding, advancing, and retreating. Walking, it could be held, is the matter of walks. Yet that is so, as it seems obvious, only metaphorically or analogically. The logical analogy is, of course, imperfect: the gold that once composed a ring can be used again, but not the laughter that composed a laugh.

It is substances alone which are literally or primarily material. Something like the thing/stuff distinction, and so something like materiality, can be extended analogically to other categories in the process of nominalization. In nominalization, we as it were create or construct entities. As the eighteenth-century writer on language, James Harris,[7] put it, we "convert even attributes into substances"

---

[7] *Hermes* (Menston: Scolar, 1969).

(*ibid.*, p. 38), and in doing so we have a choice between quasi things and quasi stuffs. It is often convenient to derive both sorts of noun from the same predicate, as it is convenient both to speak of the unbounded activity of *walking*, and to impose bounds on stretches of that activity, *walks*. So the distinction between "count nouns" and "mass terms," as it applies to such abstract beings as walking and walks, needs to be explained by reference to the less general paradigm, the distinction between substantial things and stuffs. The order of explanation here is one aspect of the primacy of substance. Unless there were natural individuals and stuffs, there could be no constructed ones.

The traditional view that individual substances are distinguished at any one time by their matter can be expressed in two equivalent ways. First, as the principle that two distinct individual substances cannot be composed of the same matter at the same time. Second, as the principle that, if two individual substances have at any time all parts in common, they are identical. Modern philosophers often deny the former and radically reinterpret the latter by introducing the paradoxical notion that material objects have temporal parts, i.e., by denying the principle that substances "exist all at once." So much the worse, I would suggest, for modern philosophy.

IV DURATION

On the traditional view, as for common sense and even for some modern philosophers, a substantial thing exists all at once, occupying space but not time. It persists through time and change, gaining and losing material parts. An event or process, on the other hand, unless strictly momentary (if such a thing is possible), takes time to exist. It does not undergo change but, as it were, unfolds. It has successive temporal parts which cannot, of course, be gained or lost.

Nevertheless, it may be that today more philosophers than not hold that there is nothing seriously mistaken, even if something is linguistically unusual, in treating things as four-dimensional time sausages with temporal parts, "stages," or "phases." A primary motive for the modern paradox has no doubt been the attraction of a logic supposed more consonant with the theory of relativity, but for many the notion of a stage is justified by its value to the theory of identity. There are some dissentient voices, however, and Wiggins rightly finds a "terrible absurdity" in the claim that a material object is a long event.[8] Yet he does not pause to justify his judgment, for example by exploring the ways in which the absurdity infects views about identity. Another way of justifying it is to draw attention to the mutually explanatory relationship between the ***endurance*** of substances and other properties of the category, other grounds of the distinction between substances and events. For example, it goes without saying that the peculiar capacity to lose old parts and gain new ones is a function of the literal ***materiality*** of substances.

---

[8] *Sameness and Substance*, pp. 25f

V. ACTIVITY

Substances are accorded a special role by the tradition in causal explanation. They are not, as we have seen, the only familiar or pretheoretical entities with "natures" to which we can appeal for the purposes of explanation. An occurrence might well be explained by reference to the nature of some mode or other, e.g., measles or corrosion. Nevertheless, the "nature" of gold or a horse has a different status from that of the "nature" of a disease or chemical process. Here there is an intuitive connection both with the materiality and the endurance of substance. For in them lies the possibility of natures that endure ***through*** the changes and interactions that they explain, natures embodied in the structure of constituent matter.

It is to make effectively the same point to say that substances possess *powers* as modes do not. We can, it is true, make sense of certain sorts of ascription of power to nonsubstances. A voice, for example, might have the power to instill fear; or a sound, the capacity to break wine glasses. Yet a voice or sound, in this context, is not something fully particular, like a piece of gold or a particular utterance. A particular speech episode either frightens or not, and there seems little point in attributing to an episode an unexercised potentiality simply on the grounds that in different circumstances it would have had a different effect. Better candidates for the status of subject of natural dispositions and potentialities can be drawn from just those modes which are in other ways logically most analogous to substances: waves, storms, fire, and the like. Yet, however that may be, it seems certain that in order to suppose a full-blooded power it is necessary to suppose an enduring thing with a (possibly unknown) structure, capable, in virtue of that structure, of affecting other things in an indefinite variety of ways and of suffering an indefinite variety of changes itself. And that is to suppose a substantial thing.

The attribution of powers and causal properties to a substance is, as it is orthodox enough to suppose, close in its character to hypothetical assertion. We can correspondingly associate the reluctance of particular modes to support powers with the fragility of their identity under the stress of hypothesis. Selecting from the past an individual man, for example Napoleon, we can intelligibly say that, if circumstances had been different at some period or moment of his childhood, he would have lived out his life in Corsica as a deformed idiot, or that he would have died a boy. The supposition does not in any way call into question the identity of the man in the "possible world" of which we speak, or frustrate our intention to refer to Napoleon. There are no constraints except the limits of physical possibility upon our supposing that an individual substance, a thing, or a particular quantity of stuff could have enjoyed a different history and yet have remained the same individual. Yet if we claim that a certain action by Napoleon at an early stage of the battle of Waterloo would have led to an immediate retreat of the allies, or to a battle with a different course and outcome, then it seems that we have in each case imagined a possible world in which the individual event, the Battle of Waterloo, simply (or at least arguably) did not occur at all. In both suppositions, only a part of the actual battle, the opening skirmish, occurred. In the second, a numerically *different* "Battle of Waterloo" occurred.

In other words, the logic that seems to make clearest sense of such suppositions about events or processes involves the divisibility of processes into temporal parts, taking the supposition to draw a line through the whole process at a point in time. The earlier part exists, whereas the later part does not exist, in the supposed possible world. The principle that two events are identical if and only if they have all parts in common thus holds between possible worlds, as we might expect from the impossibility that temporal parts should be replaced. It may seem extravagant to maintain this principle to the extent of claiming that, if we suppose the slightest divergence from the actual course of an event, then we suppose a different event. Yet it is difficult to see any logically pure ground for stopping short of that claim. Even if we are seldom so strictly logical in ordinary speech, the point of contrast with substances is clear. In the case of substances, as it is widely recognized, there are comparable restrictions on the supposition that an individual might have had different origins. We cannot intelligibly speculate, for example, that in certain circumstances Napoleon would have been the child of different parents. But there are no such restrictions relating to the *consequent* life history of a substance. There is nothing in the least problematic in respect of identity about the claim, 'If it had been given proper conditions, that pickled egg would by now have been a fine hen'. All this is involved in the point that a substance *endures* rather than *lasts*. It is, moreover, the condition of the explanatory role of substances as the subjects of powers, a role celebrated in their different ways by both Aristotle and Locke. We can identify all at once a potential subject of indefinitely various and perhaps, for us, unpredictable change: a subject to be actively tested and explored, or to have its hitherto unknown "normal" life history, with all its metamorphoses, however exotic and unexpected, observed and recorded.

One objection to according substances their traditional pride of place takes the form of an argument that the explanation schema embodied in the notion of a substance is outmoded. When in the mechanist youth of modern science the notion was reinterpreted, the essences of such things as horses and gold were no longer thought to be ultimates. Yet the new ultimate, matter with an essence, seemed to fall squarely enough into the category of substance, no doubt because it was tailored to do so. The ultimates of present-day physical theory, however, seem less accommodating to traditional metaphysics. Electrons, photons, and the like are not helpfully conceived of as shaped space occupants composed of a uniform superstuff. Quantum mechanics, as much as the theory of relativity, has fostered the tendency to underplay the logical differences between things and events. Physics can seem to have shown that material things are really probabilistic processes.

If the entities postulated by physics cannot be straightforwardly categorized as "substances" or "material objects" (i.e., if the standard properties of the category of substance do not all pertain to them), that is not, on reflection, surprising. It is at any rate consonant with a distinction which can be advocated on epistemological grounds, but which conceptualists are almost bound to reject. That is the distinction between the pretheoretical level of experience and the level of speculative theory about the fundamental nature of the objects of experience. Given that distinction, then it is very intelligible that the interlocking features of the category of substance, especially in so far as they involve the natural and real individuality of substances, reflect and signal the impingement of reality upon us (unmediated by any "concept") at the level of *experience*. At what other level *could* there be direct cognitive contact with reality? *Nature, thing, structure, power, stuff*—all these notions function in their relation to the category of substance in such a way as to mark what is properly assumed to be open to a full theoretical account or explanation, and to mark it in a way that is neutral as between rival theories precisely because this, their primary employment, is at the pretheoretical level of experience. If, as seems inevitable, the physical theory employed in that further explanation involves the postulation of further individuals beyond the limits of direct experience (such as the primary particles of modern physics), then it is clearly impossible that we should identify those individuals, as we can identify complex and sensible material objects, without knowing what kinds of things they are, and without having some systematic knowledge of their nature. For theoretical entities are individuated *through* our speculation as to their nature and kind, as perceived material objects are not.[9] It is therefore unsurprising that such objects do not fall into the category of substance as traditionally described. So far is this from being an argument against realism, it is just what the realist should expect; which is not to say that there is nothing problematic about the ontology of quantum physics, or that there are *no* constraints on what makes sense in a description of reality at that level.

This issue surely identifies an important source of the crisis of confidence in the category of substance which led to its present disrepute. In Aristotelian philosophy, the theory of substance was developed in the context of the assumption that, for the most part, scientific explanation and theory could effectively begin and end with the natures of the relatively large and complex individuals, above all biological individuals, which we can pick out prior to theory at the level of experience. When it came to be felt that a science based on that assumption left too much unexplained, the mechanical hypothesis offered an alternative model that still held the levels of experience and theory together, although in a different way; for essentially the same mechanical concepts and the same category of a *stuff* as we employ at the level of experience were employed in a theory about the imperceptibly small. Indeed, it was evidently felt necessary to show that matter subject to mechanical modifications fitted the traditional philosophical conception of substance even better than the sensible material objects in relation to which that conception had originally been constructed. Other felt needs, relating to God and the soul, played a distorting role in the debate. But the real crisis came with the fall of mechanism and the sense that the traditional conception of substance is inappropriate, at least as it stands, for dealing with the postulated entities of physics. In effect, the theory of substance was split between the philosophy of physics and the philosophy of experience (or of the way we think about the world at the level of experience). It is not, perhaps, surprising if some have felt that the second fragment is a disreputable and reactionary relative of the first, a relic of outdated worldviews. Yet that judgment is insensitive to the possibility of, and need for, a philosophical explanation of the structure of our thinking and, in particular, of the contribution made to that structure by reality itself, not through its tendency to confirm or disconfirm the stories scientists tell about it, but by its impingement on us at the pretheoretical level at which we make our first and most fundamental discriminations.

---

[9] There is a familiar question how far it is possible to identify individuals postulated by one theory by those postulated by a different theory, e g., Newton's atoms with Einstein's. If it is admitted to be possible, then it must be supposed possible to identify or individuate theoretical objects while having only approximate knowledge of their nature. Nevertheless, identifying objects through an approximately correct theory is different from individuating, without knowing what it is, a particular object picked up in the park.

VI. MEMBERSHIP OF A NATURAL KIND

The still generally accepted principle that every individual substance belongs to a universal kind or sort has had, in its history, two very different sources of support. First, together with the view that the only natural and real species and genera are the species and genera of substances, the principle constituted part of the metaphysical conception of individual substances as embodied forms. The second grounding of the principle, on the other hand, has been the conceptualist theory of identity which began to be formulated by seventeenth-century opponents of Aristotelianism, most notably by Locke. The universal kind is held to be essential to the individual substance just because a universal idea or concept must come before individuation: individuation is always relative to a sortal concept.

In its standard modern form, however, the conceptualist theory takes *automobile* or even *piece of stone* to be as good a sortal concept as *horse*. If it is thought that all individuation is a matter of our slicing up the world by means of concepts, there might well seem no basis for the careful Aristotelian distinctions between the natural and real unity of a horse, the accidental unity of a piece of stone or a dead body, and the artificial unity of an axe or house. On the other hand, if we are neither Aristotelians nor conceptualists, it might seem that a lump of lead is quite an impressive individual in its own right, perhaps no less so than a plant; and that artifacts, while sometimes physically unitary, like most teaspoons, are sometimes hardly more than notionally so, like a steam engine. In that case, since "lumps" or "pieces" do not form a kind in any significant sense, the principle that all individual substances belong to kinds would seem dispensable.

The converse principle, however, that the only real or natural species are species of substances, still has much to be said for it, some of which has been said above under the title "unity."

VII CONCLUSION

It is to be hoped that the above discussion of six properties traditionally ascribed to the category of substance has presented the theory of substance (or something distilled from theories of substance) as more plausible, more unitary and coherent, and more capable of explaining the actual structure of our thought and its relation to reality than common prejudice allows. The question of the basis of the natural individuality or unity of paradigm substances has not been answered explicitly, but there have been heavy enough hints of the form that an answer might take. The most promising explanation of that unity lies with a further connection between the various conditions of substantiality. Substances, and only substances, are material, and only substances can be coherent. Coherence is a causal relation that presupposes materiality. Only substances, therefore, can have the kind of given physical boundary that a coherent material thing can have. Only substances can be *materially* unitary and discrete objects. This somewhat obvious proposal has not been popular, but it does have its antecedents in the arguments of those ancient and early-modern atomists for whom, in effect, the continued existence of a complex individual substance just is the continued existence of a coherent mass of particles or matter. For such an answer to be adequate, however, it is necessary to say more about what constitutes a natural individual's "coherence." It is implausible to suggest that a cannon ball and the piece of chewing gum sticking to it together form such an individual. What matters is not just coherence, but also the nature and causality of the coherence.

A theory of substantial identity has these and other questions to answer. It is submitted that the traditional doctrines discussed above suggest lines of approach likely to prove very much more explanatory than current orthodoxies. They should also help us to understand what is wrong with the sophistical, paradoxical arguments on which conceptualism and the doctrine of ontological relativity are grounded.

MICHAEL AYERS

Wadham College, Oxford

THE INDETERMINACY THESIS REFORMULATED*

IN *Word and Object*, W. O. Quine[1] presents the following formulation of his thesis of the indeterminacy of translation:

> There can be no doubt that rival systems of analytical hypotheses can fit the totality of speech behavior to perfection, and can fit the totality of dispositions to speech behavior as well, and still specify mutually incompatible translations of countless sentences insusceptible of independent control (*ibid.*, p. 72).

I. INDETERMINACY, ANALYTICAL HYPOTHESES, AND DICTIONARIES

To simplify matters, I shall here assume that systems of analytical hypotheses contain only analytical hypotheses in their simple form,[2] that is, hypotheses that equate a word of one language (the language being translated) with a word or open sentence of another language (the home language). I shall also assume that analytical hypotheses are designed to translate one individual's idiom, or speech, into another's. Thus, for my purposes, a system of analytical hypotheses is a dictionary, or translation manual, correlating the words of one individual's idiom with the words of another's. What exactly is involved in systems of analytical hypotheses fitting an individual's speech behavior and his dispositions to behavior will not concern me here. For my purposes, it is sufficient to note that any two systems that equally fit an individual's speech behavior and his dispositions to behavior will, by Quine's lights, be empirically equivalent manuals.

Hereafter I shall speak simply of empirically equivalent rival systems of analytical hypotheses, or empirically equivalent rival dictionaries fitting a person's speech, or, more simply still, of rival (empirically equivalent) manuals. The indeterminacy thesis is then, or at least entails, the following claim:

(C.I.T.) There can be rival (empirically equivalent) manuals fitting a person's speech.

I say 'at least entails' because, although it is clear that every case of indeterminacy of translation involves a case of rival dictionaries, it is not clear that every case of rival dictionaries constitutes a case of

---

* Thanks are due to Robert Brandom, Joseph Camp, Jr., Clark Glymour, John Haugeland, Gerald Massey, and an unnamed referee from this JOURNAL for comments on various drafts

[1] Cambridge: MIT, 1960.

[2] The notion of "analytical hypotheses in their simple form" is Quine's, *ibid.*, p 70

indeterminacy of translation. For instance, there may be a pair of rival dictionaries which give different translations for some words of an individual's speech but which give logically equivalent translations for all full sentences of the individual's speech. (For an example of such manuals cf. section IV below.) Such a pair of rival manuals may, by Quine's lights, constitute an example of the inscrutability of reference but not an example of the indeterminacy of translation. For indeterminacy, one needs incompatible translations of whole sentences.

The above formulation of the indeterminacy thesis, or, at least, of a Consequence of the Indeterminacy Thesis [hence the label (C.I.T.)], is very close to the following formulation in Quine's *Word and Object:*

> The thesis is then this: manuals for translating one language into another can be set up in divergent ways, all compatible with the totality of speech dispositions, yet incompatible with one another (*ibid.*, p. 27).

II. RIVAL DICTIONARIES VERSUS DIFFERENT DICTIONARIES

The claim that there can be merely different dictionaries that fit an individual's speech is surely uncontroversial. Thus, a dictionary that fits my speech behavior, constructed by an Italian, will, presumably, include the entry: book/libro; whereas a dictionary constructed to the same effect by a German will include the entry: book/Buch (hereafter case 1). Now, if we were to count such *merely different* dictionaries as *rival* dictionaries then (C.I.T.) would be trivially true and not the least bit controversial. Of course, (C.I.T.) is far from trivially true and uncontroversial. If this is so, there must be some interesting distinction between genuine rival dictionaries and merely different dictionaries. Indeed, if (C.I.T.) is not trivially true, then there must be some interesting distinction between empirically equivalent rival dictionaries and merely empirically equivalent different dictionaries. The rest of this paper will be largely devoted to showing that Quine's own philosophy of language is incompatible with the claim that such a distinction exists.

Before considering a Quinean attempt to separate the notions of rival and different empirically equivalent dictionaries, let us consider some intuitive ways of separating these two notions.

In the above mentioned case of the Italian and German dictionaries, the dictionaries are intuitively not genuine rival dictionaries, because the terms 'libro' and 'Buch' respectively equated by the two dictionaries with my term 'book' are terms of different languages, and are themselves most plausibly equated with each other. This consideration suggests the following criterion for rival dictionaries:

(C) Two empirically equivalent dictionaries fitting an individual's speech are rival dictionaries, if and only if they can be translated into one and the same language yielding at least one conflict over the interpretation of some word of the individual's speech.

Now, we need not, thankfully, bother attempting to explicate the difficult notion of "conflict of interpretation over some word," for there are at least three reasons why (C) is not acceptable to a Quinean.

First, (C) relies on the intuitive notion of "same language." This notion, like its sister notion of "analyticity," is rejected by Quine. This alone is enough to make (C), unsupplemented by a Quinean notion of "same language," unacceptable to Quineans. Later we shall see that, even given a Quinean notion of same language, we cannot construct an adequate Quinean notion of rival manuals (cf. section V below).

Second, (C) makes use of the notion of "translation." Since this notion is, in its normal usage, closely related to the notion of "preservation of meaning" (where meaning is not simply a matter of Quinean stimulus meaning), it cannot be uncritically used by Quineans.

Third, (C), combined with Quine's claims about the indeterminacy of translation, has the consequence that any two dictionaries fitting an individual's speech count as rival dictionaries. Thus, consider the following scenario (hereafter case 2).

A German and French linguist, Moritz and Pierre, are attempting radical translation with regard to the speech of some native. Pierre's dictionary contains the entry: (1) gavagai/lapin; and Moritz's contains the entry: (2) gavagai/Kaninchen. Now, if the indeterminacy thesis is correct, the following may be an adequate translation of Pierre's (1) into my idiom: gavagai/undetached rabbit part. After all, if there is no fact of the matter as to whether the native's term 'gavagai' means 'rabbit' rather than 'undetached rabbit part', then, by the same token, there is no fact of the matter as to whether Pierre's term 'lapin' means 'rabbit' rather than 'undetached rabbit part'. At the same time, an adequate translation of Moritz's (2) into my idiom may be: gavagai/rabbit. In other words, by (C), Moritz and Pierre have rival manuals for the native's speech!

The same point may be made more starkly, albeit more confusingly, by considering a case involving two speakers of the same language (hereafter case 3). Suppose two German linguists, Karl and Otto, are attempting radical translation with regard to the speech of our native. Karl's dictionary contains the entry: (3) gavagai/Kaninchen; and Otto's dictionary contains the entry: (4) gavagai/Kaninchen. Now, if the indeterminacy thesis is correct, the following may be an adequate translation of Karl's (3) into my idiom: gavagai/undetached rabbit part. After all, if there is no fact of the matter as to whether the native's term 'gavagai' means 'rabbit' rather than 'undetached rabbit part', then, by the same token, there is no fact of the matter as to whether Karl's term 'Kaninchen' means 'rabbit' rather than 'undetached rabbit part'. At the same time, an adequate translation of Otto's (4) into my idiom may be: gavagai/rabbit. In other words, by (C), the two Germans have rival manuals for the native's speech, even though they have identical manuals!

More generally, suppose that both *A*'s term *X* and *B*'s homophonically equivalent term *X* may, by Quine's lights, be translated into either of terms *Y* or *Z* of *C*'s idiom. Then in any case where *A* translates some individual *D*'s term *W* into his term *X* and *B* likewise translates *W* into his term *X*, then, if (C) is correct and *A* and *B*'s manual both fit *D*'s speech, this constitutes a case of rival dictionaries for *D*'s speech!

In other words, (C), combined with Quine's claims about the indeterminacy of translation, entails the claim that any pair of (empirically equivalent) translation manuals, including a pair of identical manuals, constitutes a case of rival manuals!

Clearly, (C) is unacceptable, or, at least, unacceptable to any Quinean.

III RIVAL DICTIONARIES: A QUINEAN ACCOUNT AND ITS PROBLEMS

No doubt, if we put our minds to it, we could concoct other suggestions for explicating the notion of rival (empirically equivalent) translation manuals. We need not bother putting our minds to it, however, since Quine himself provides material from which such a suggestion may be drawn.

In his article "On Empirically Equivalent Systems of the World,"[3] Quine says:

> So I propose to individuate theories thus: Two formulations express the same theory if they are empirically equivalent and there is a reconstrual of predicates that transforms one theory into a logical equivalent of the other (*ibid.*, p. 320).

Presumably, Quine is also here committing himself to the claim that, if two theories are either not empirically equivalent or cannot be transformed into logical equivalents of each other through the reconstrual of predicates, then they are not the same theory.

Now, dictionaries or systems of analytical hypotheses are just theories about what the terms of a given individual's (the translatee's) idiom mean. And, as assumed earlier, any two dictionaries that fit to perfection an individual's speech and his dispositions to speech are, at least by Quine's lights, empirically equivalent. Thus, we have the following Quinean criterion for individuating dictionaries:

---

[3] *Erkenntnis*, IX (1975): 313–328.

(Q.C.1) Two dictionaries, both of which fit an individual's speech, are rival dictionaries if and only if there is no reconstrual of words which transforms one into a logical equivalent of the other.

As regards case 1 above, the case of the Italian's and German's different dictionaries for my speech, (Q.C.1) yields the intuitively correct result that this is not a case of genuine rival dictionaries. By reconstruing the German's term 'Buch' as the Italian's term 'libro' and analogously reconstruing all the German words in the German's dictionary, we end up with two identical, hence logically equivalent, dictionaries.

Similarly, by (Q.C.1), the German linguists Karl and Otto, in case 3 above, do not have rival dictionaries for the native's speech. By "reconstruing" Karl's term 'Kaninchen' as Otto's term 'Kaninchen' and similarly "reconstruing" the words of Karl's dictionary, we arrive at two identical, and hence logically equivalent dictionaries.

Unfortunately, (Q.C.1) is not consistent with Quine's own examples of indeterminacy of translation.

Let us consider a typical Quinean case (hereafter case 4). Linguists Bob and Barry are attempting radical translation of some native's speech. Bob's dictionary includes the entry: gavagai/rabbit; whereas Barry's dictionary includes the entry: gavagai/undetached rabbit part. By reconstruing Bob's term 'rabbit' as Barry's term 'undetached rabbit part', and likewise reconstruing other terms where necessary, we arrive at two identical, hence logically equivalent, dictionaries. Thus, by (Q.C.1), Bob and Barry do not have rival dictionaries for the native's speech!

IV. (Q.C.1): SOME FURTHER CONSEQUENCES

We have seen that, judged by the standard of our Quinean account (Q.C.1) of the notion of rival dictionaries, Quine's famous gavagai-rabbit-undetached-rabbit-part example collapses to a case of merely different, and not rival, dictionaries. Yet what of other purported examples of rival dictionaries?

Consider the following variant of a case presented by Gerald Massey[4] (hereafter case 5). Native *N* speaks a first-order quantificational, "English-type" language (without singular terms). Further, *N* always prefaces his utterances with either the term 'Yeg' or 'Neg'. Two linguists, Joe and John, are engaged in the project of translating *N*'s speech. Joe treats *N*'s term 'Yeg' as the affirmative modifier 'It is the case that' and *N*'s term 'Neg' as the negative modifier 'It is not the case that'. Otherwise, Joe maps the terms of *N*'s speech homophonically. Thus, Joe maps *N*'s term 'rabbit' onto his own term 'rabbit', and maps *N*'s term 'or' onto his own term 'or', etc. John, however, treats *N*'s term 'Yeg' as the negative modifier 'It is not the case that' and *N*'s term 'Neg' as the affirmative modifier 'It is the case that'. Further, John maps *N*'s general terms onto their "homophonic complements." For instance, he maps *N*'s term 'rabbit' onto his (John's) term 'nonrabbit'. Further, John maps the logical particles of *N*'s speech onto their duals, 'v' onto '.', etc., and he maps *N*'s quantifier '(E*x*)' onto his (John's) quantifier '(*x*)' and *N*'s quantifier '(*x*)' onto his (John's) quantifier '(E*x*)'. Thus, where *N* utters the sentence 'Yeg; (E*x*) (Rabbit *x*)', Joe translates it as 'It is the case that; (E*x*) (Rabbit *x*)', while John translates it as 'It is not the case that; (x) (nonrabbit *x*)'.

---

[4] Cf. "Indeterminacy, Inscrutability, and Ontological Relativity," in *Studies in Ontology*, N. Rescher, ed., *American Philosophical Quarterly*, Monograph Series, XII (1978).

By Quine's lights, this would constitute an example of the inscrutability of reference and not an example of the indeterminacy of translation. Although Joe and John have rival theses about the reference of all the native's general terms, they give logically equivalent translations for all of the native's utterances.

Yet, according to (Q.C.1), Joe and John do not here have rival dictionaries for *N*'s speech, since there is a simple reconstrual of words (Joe's affirmative modifier 'It is the case that' onto John's negative modifier 'It is not the case that', Joe's general term 'rabbit' onto John's term 'Nonrabbit', Joe's quantifier '(E*x*)' onto John's quantifier '(*x*)', etc.) which renders the two manuals identical and hence logically equivalent. So, according to (Q.C.1), this does not constitute a case of the inscrutability of reference.

More generally, any two dictionaries for an individual's speech may be transformed into each other by reconstrual of words. Consider any two such dictionaries *F* and *G*, both of which fit individual *D*'s speech. *Ex hypothesi,* there is a mapping of the words in *G* onto the words of *D*'s speech (by the symmetricality of the mapping function). Similarly, there is a mapping of the words in *F* onto the terms of words of *D*'s speech. So, by the transitivity of the mapping function, there is a mapping of the terms of dictionaries *F* and *G* onto each other.

In other words, (Q.C.1) is incompatible with (C.I.T.), since it has the consequence that there are never two rival dictionaries fitting an individual's speech.

V. SUMMARY AND VARIOUS OBJECTIONS

Basically the argument so far may be summarized as follows:

(1) If the indeterminacy thesis or the inscrutability of reference thesis is true, then there are rival empirically equivalent manuals of translation.
(2) If Quine's claims about theory identity are true, then if two empirically equivalent manuals are in fact rival manuals then they must not be reconcilable by word reconstrual.
(3) Yet any two empirically equivalent manuals can be reconciled by word reconstrual.

Therefore,

(4) Either the indeterminacy thesis and the inscrutability thesis are false, or Quine's claims about the identity of theories are false.

Note, a Quinean cannot simply solve this dilemma by accepting the second disjunct of (4)—such a move would leave him in the position of owing an account of how empirically equivalent translation manuals can be rival manuals. Indeed, later (cf. section IX we shall see that the indeterminacy thesis conflicts with other tenets of Quine's philosophy, in particular, his physicalism.

Before attempting to develop the argument further, I want to consider various objections that were elicited by an earlier, unpublished version of this paper. This will, I hope, help clarify the argument.

Objection 1: Since, in case 4 for example, Bob and Barry are speakers of the same language, we cannot simply reconstrue Bob's words in order to make their dictionaries identical. Hence the use of (Q.C.1) to show that the above cases do not represent cases of the indeterminacy of translation and the inscrutability of reference is illegitimate.

Answer: Now, we may indeed balk at the above use of (Q.C.1) to "reconcile" the prima facie rival translations involved in the above mentioned cases. Yet where this intuitive balk exists it is the result of the tacit, and totally "un-Quinean" assumption that the linguists involved are speakers of the same language, ***in the ordinary sense of that notion.*** Given this assumption, it would be illegitimate, for instance, in case 4 above to take Bob's term 'rabbit' as being equivalent to (as having the same meaning as) Barry's term 'undetached rabbit part'. Yet this assumption cannot simply be accepted into the Quinean framework. Before Quineans could explicitly accept such an assumption they would need to give some Quinean account of the *ordinary* notion of being speakers of the same language. Yet any such account would need to steer clear of the conventional notion that two individuals are speakers of the same language if they utilize the same terms to convey *the same meanings*.

Indeed, there is a strictly Quinean notion of being speakers of the same language, namely, that of being speakers who can converse freely, or at least without any irreparable breakdown, using only homophonic manuals of translation. This notion, unlike the ordinary notion of being speakers of the same language, does not carry with it, however, the implication that, if *X* and *Y* are speakers of the same language, then the only legitimate mapping of one's terms onto another's is the homophonic mapping.

For Quine, the claim that Bob and Barry are speakers of the same language does not preclude our construing Bob's term 'rabbit' as being equivalent to Barry's term 'undetached rabbit part'. Thus, Quine says that the indeterminacy thesis begins at home.

Objection 2: Does the above usage of Quine's claims about theoretical equivalence not go beyond mere predicate reconstrual?

Answer: It is true that, in speaking of equivalent theories, Quine only specifically talks of reconstrual of *predicates*, whereas I, in dealing with equivalence of translation manuals, spoke of reconstrual of *words*. This does raise the question of whether this move is legitimate.

Now, clearly Quine would have no objection to reconstruing names as well as predicates in order to show that two theory formulations are in fact formulations of the same theory. Thus, consider a theory *T* ranking heavyweight boxers and an empirically equivalent theory *T*1 that contains the same sentences as *T*, save that wherever the name 'Muhammad Ali' occurs in *T* the name 'Cassius Clay' occurs in *T*1. Surely, Quine would allow that *T* and *T*1 are in fact formulations of the same theory, since they may be rendered logically equivalent by reconstruing *T*'s name 'Muhammad Ali' as 'Cassius Clay'.

Yet note, in dealing with the Massey-type purported rival manuals in case 5 above, we went beyond reconstrual of predicates and names to reconstrual of quantifiers. To reconcile Joe's and John's dictionaries for *N*'s speech, we reconstrued, among other things, Joe's quantifier '(E*x*)' as John's quantifier '(x)'. A Quinean might here object that there is a fact of the matter about whether a particular expression *e* of a person's idiom is a universal quantifier or an existential quantifier, and since Quine's criterion of equivalence of theories does not explicitly allow reconstrual of quantifiers, we cannot reconstrue Joe's universal quantifier '(x)' as John's existential quantifier '(E*x*)' in order to show that Joe and John have equivalent dictionaries for *N*'s speech.

Yet, how can this objection be maintained when, as in our case, there are two manuals for translating Joe's idiom into John's idiom, both of which fit Joe's speech dispositions, though one translates Joe's expression '(x)' as John's universal quantifier and the other translates it as John's existential quantifier? (For the purposes of constructing a *reductio,* we take for granted the claim that John's expression '(x)' is a universal quantifier and his expression '(Ex)' is an existential quantifier.) If there is no meaning except that which is fixed by speech dispositions then there is no fact of the matter as to which of such manuals is (exclusively) correct. In other words, in our case there is no fact of the matter whether Joe's expression '(Ex)' is a universal or an existential quantifier. The indeterminacy thesis itself conflicts with the objector's hypothesis that there is a fact of the matter as to whether Joe's expression '(Ex)' is an existential rather than a universal quantifier.

We shall now consider an objection to our argument for the conclusion that Quine's indeterminacy thesis is incompatible with his notion of the equivalence of theories which will lead to a new criterion of rival dictionaries.

VI. THE REJECTION OF (Q.C.1) IN FAVOR OF (Q.C.2)

In developing (Q.C.1) from Quine's remarks about the identity of theories, I have been tacitly assuming that what Quine explicitly says about the identity of *world* theories—for this is the explicit subject matter of Quine's "On Empirically Equivalent Systems of the World"—goes also for those less global theories that constitute translation manuals, or dictionaries. This assumption may raise the objection that we have erred in considering translation manuals *in vacuo,* that what we should consider, if we are to make our case, is manuals in the context of total world theories.

Let us consider how this claim might be developed as an objection to our argument.

According to Quine, two theorists may share exactly the same theory of the world, indeed they may assent to the same set of sentences (any member of that set having the same stimulus conditions for both theorists), except that they disagree about the translation of some individual's speech. Let Mork and Mindy be two such theorists. Aside from offering different dictionaries for *N*'s speech, they affirm the same set of sentences (with the same stimulus conditions). Mork's dictionary for *N*'s speech (hereafter Mork's dictionary) includes the following entry: (A1) gavagai/rabbit. Mindy's dictionary for *N*'s speech (hereafter Mindy's dictionary) includes the entry: (B1) gavagai/undetached rabbit part. Let us also suppose that Mork's total world theory contains the claims: (A2) Rabbits are not identical to undetached rabbit parts; and (A3) Rabbits are identical to rabbits. Then, *ex hypothesi,* Mindy's total world theory will contain the claim: (B2) Rabbits are not identical to undetached rabbit parts.

In this case, if we try to reconcile Mork's and Mindy's dictionaries by simply reconstruing Mindy's term 'undetached rabbit part' as (being equivalent to) Mork's term 'rabbit', we render Mindy's total world theory no longer compatible with Mork's. For, in this case, Mindy's (B2) becomes: (B2*) Rabbits are not identical to rabbits—which is inconsistent with Mork's (A3).

So far the objections seems good. It seems that, in the above case, we cannot reconcile the two dictionaries without contradicting our hypothesis that Mork and Mindy have, aside from their dictionaries, the same world theories. Thus, it seems we have here a case of genuine rival dictionaries and hence a case of genuine indeterminacy of translation.

The point is that we cannot simply use (Q.C.1) to reconcile two translators' different dictionaries for an individual's speech, where such a reconciliation would lead to the creation of contradictions between the translators' background theories. Our criterion for identity of dictionaries (Q.C.1) is not acceptable in its current form. The above considerations suggest an obvious amendment, however; thus we have the new criterion:

> (Q.C.2) Two dictionaries, both of which fit an individual's speech, are rival dictionaries if and only if there is no reconstrual of words that transforms one into a logical equivalent of the other without also creating some (new) contradiction between the respective background theories in which the respective dictionaries are embedded.

VII. (Q.C.2), (I.T.), AND BEYOND

In the above case, our attempted reconciliation failed because we invariably mapped Mindy's term 'undetached rabbit part' onto Mork's term 'rabbit'. Yet note, if we treated Mindy's term 'undetached rabbit part' *as it occurs in her dictionary* as a technical term to be mapped onto Mork's term 'rabbit', while mapping other occurrences of the term 'undetached rabbit part' in Mindy's world theory onto Mork's term 'undetached rabbit part', then both Mindy's dictionary and her total world theory are identical with, and hence logically equivalent to, Mork's dictionary and total world theory. Hence, by our criterion (Q.C.2), Mork and Mindy do not have rival dictionaries.[5]

---

[5] Of course, this "double strategy" may well have to be taken with some of Mindy's other terms. For example, the term 'identical' may have to be given differ-

Now, it might here be objected that this gimmick of only partially reconstruing occurrences of Mindy's term 'undetached rabbit part' is, by Quine's lights, illicit. It depends on the assumption that that term as it occurs in Mindy's dictionary has a different meaning from that which it has in occurrences in other parts of Mindy's world theory, though the term as it occurs in Mindy's dictionary is stimulus synonymous with its occurrences in the nondictionary part of Mindy's world theory. For Quine, there is no fact of the matter as to whether it has such a "double life."

To see this, consider the case of linguists *X* and *Y*, who are engaged in the project of translating Mindy's world theory into their own idioms. *X* translates Mindy's world theory homophonically into his language, *Y* translates it homophonically save that occurrences of the term 'undetached rabbit part' in Mindy's dictionary for *N*'s speech are mapped onto *Y*'s term 'rabbit'. Clearly, both these manuals would equally fit Mindy's dispositions to speech. Thus, from a Quinean point of view, there is no fact of the matter as to which manual is correct. In other words, since *Y*'s manual treats Mindy's term 'undetached rabbit part' as having a double life and *X*'s does not, there is no fact of the matter whether Mindy's term does or does not have such a double life.

Yet note: the case of linguists Mork and Mindy only represents a case of rival translation manuals if Mindy's term 'undetached rabbit part' is uniform in meaning throughout her world theory. Thus, if there is no fact of the matter as to whether there is such uniformity, then there is no fact of the matter as to whether Mork and Mindy have genuine rival manuals. That is to say, there is no fact of the matter as to whether this case is one of the indeterminacy of translation.

Indeed, when we reconsider our other purported cases of indeterminacy of translation and inscrutability of reference, we see that the same conclusion may be drawn: there is, by Quinean lights, no fact of the matter as to whether there is any indeterminacy of translation.

Thus, consider the case of linguists Joe and John in case 4 above, who respectively translate *N*'s utterance 'Yeg; (E*x*) (Rabbit *x*)' as 'It is the case that; (E*x*) (Rabbit *x*)' and 'It is not the case that; (*x*) (nonrabbit *x*)'. Now, if there is no fact of the matter as to whether *N*'s terms 'Yeg', 'rabbit', and '(E*x*)' mean, respectively, 'It is the case that', 'rabbit', and '(E*x*)' as opposed to 'It is not the case that', 'nonrabbit', and '(x)', then there is no fact of the matter as to

---

ent construals according to whether it occurs inside or outside her dictionary. I owe this observation to a referee from this JOURNAL.

whether Joe's terms 'It is the case that', 'rabbit', and '(Ex)' mean, respectively, 'It is the case that', 'rabbit', and '(Ex)' rather than 'It is not the case that', 'nonrabbit', and '(x)'. That is to say, there is no fact of the matter as to whether Joe and John have merely different dictionaries or genuine rival dictionaries for *N*'s speech.

Put in another way: there are rival manuals for translating John's idiom into Joe's idiom, a homophonic manual and a nonhomophonic manual. Translating by one manual leads to the conclusion that Joe and John have rival manuals for *N*'s speech. Translating by another manual leads to the conclusion that Joe and John do not have rival manuals. Since, if Quine is right, there is no fact of the matter as to which manual is (uniquely) correct, then, by Quine's lights, there is no fact of the matter as to whether Joe and John have genuine rival manuals for *N*'s speech.

Thus, it seems that the theses of the indeterminacy of translation and the inscrutability of reference are inconsistent with other parts of Quine's philosophy of language. In particular, it is inconsistent with his account of the equivalence of theories. To make sense of the indeterminacy and the inscrutability theses we need to make sense of the notion of rival, though empirically equivalent translation manuals. Yet it is this very type of notion, notions that incorporate the idea that there can be real *semantic* differences (i.e., conflicts) even where there is empirical equivalence, at which Quine's attacks on "the museum myth of the mental," on analyticity, and on the determinacy of translation and the scrutability of reference are aimed. To put the point yet another way, the indeterminacy thesis entails that there can be two rival, though empirically equivalent theories of what a speaker's utterances mean. Yet if we, *like Quine*, "recognize with Pierce that the meaning of a sentence turns purely on what would count as evidence for its truth," we should conclude that any two manuals that assign the same stimulus conditions to utterances of a translatee's idiom are in fact assigning the same meaning to those utterances.

Before proceeding, I would like to register one observation and one reservation about the above argument.

First the observation. I have not yet considered the objection that I have only dealt with cases of two translators offering purported rival dictionaries for a translatee's speech. Indeed, I have not, so far, addressed (directly) the case of a single translator who offers two apparently rival dictionaries for an individual's speech. Yet from the above we may clearly draw an answer to this objection.

Suppose Joe offers two purported rival manuals, *A* and *B*, for *N*'s speech. In translating Joe's idiom onto our idiom, or, for that matter, when translating Joe's idiom onto itself, we may, if the indeterminacy thesis is correct, reconstrue Joe's words so that *A* and *B* are (intranslated form) identical. A non-Quinean may here object that to do this is to treat Joe as using one language in constructing manual *A* and using another language in constructing manual *B*. Yet, continues our hypothetical non-Quinean, there is a fact of the matter as to whether Joe has switched languages or not. If he has in fact not switched languages, then *A* and *B* are rival manuals. For this objector, there is a real fact of the matter about whether Joe, in moving from manual *A* to manual *B*, has switched theories about the meaning of *N*'s speech or merely switched languages.[6]

All this is something a thorough-going Quinean must reject. He will reject this because it relies on a thoroughly un-Quinean notion of same language. If we could make sense of such a fine notion of sameness (and hence of change) of language, we would have all the material we need for reconstructing the old analytic-synthetic distinction. A sentence would be analytic for individual *X* if and only if its rejection represents a change in *X*'s language as opposed to a change in *X*'s world theory.

So much for the objection. Now for the reservation. So far we have proceeded on the assumption that the translating linguists agree on what segments of the translatee's noise productions count as words. If two linguists disagree over this matter, then we cannot reconcile their dictionaries in the manner suggested above. So it seems we are only entitled to the conclusion that Quine cannot reconcile the claim that two linguists might agree on what segments of a translatee's noise production count as words and yet offer rival (empirically equivalent) translation manuals for the translatee's speech with his criteria of equivalence of theories. That Quine is indeed committed to this claim is evidenced by his claim that an individual's words may be permuted onto themselves in a nonhomophonic mapping that preserves all his dispositions to speech.

VIII. INDETERMINACY, MENTALISTS, AND QUINE

The central strategy I have been pursuing is that of using Quine's own notion of theoretical equivalence to show that purported cases of genuine rival translation manuals, upon which the indeterminacy thesis and the inscrutability-of-reference thesis depend, collapse into cases of merely different translation manuals. Yet perhaps we should construe the indeterminacy and inscrutability theses, not so much as theses Quine is willing to endorse outright, but rather as theses he believes those who believe in old-fashioned meanings (the idea idea, etc.) may be forced into accepting. Perhaps the indeterminacy thesis and the inscrutability-of-reference thesis are best construed as a

---

[6] The above two paragraphs arose from discussions with Haugeland.

reductio of the mentalist's conception of meaning: a mentalistic theory of meaning forces one to accept as rivals two theories (two different translation manuals) when there is no empirical difference between the two.

Under this interpretation, the apparent conflict between the indeterminacy thesis and the inscrutability-of-reference thesis and Quine's criterion of the equivalence of theories is no real conflict, since mentalists do not typically share that criterion of the equivalence of theories. For the typical mentalist, what makes two different empirically equivalent theories genuine rival theories is that they express different ideas, different conceptions, of the way the world is (in some or other aspect). For the typical mentalist, what makes *A*'s dictionary for *N*'s speech a rival to *B*'s dictionary for *N*'s speech is that it expresses a different idea about the meaning of *N*'s words.

Thus, suppose again that *N* is our native and *A*'s dictionary for *N*'s speech contains the entry 'gavagai/rabbit', whereas *B*'s dictionary contains the entry 'gavagai/undetached rabbit part'. According to our mentalist, this constitutes a case of rival dictionaries, because the terms 'rabbit' and 'undetached rabbit part' express/stand for different ideas.

On this interpretation, Quine's claim is that, given the mentalist's notions of meaning and rival dictionaries, there will in fact be rival dictionaries that are such that no empirical fact (and presumably no physical fact, e.g., a fact about the distribution of physical particles) favors one over the other. For Quine, this amounts to a refutation of the mentalist's position.

One obvious problem with this interpretation is that it does not at all gibe with the textual evidence. Quine continually presents the indeterminacy thesis and the inscrutability-of-reference thesis not as part of a greater *reductio* but as theses that he outright endorses. Thus, at one place he says:

> I am persuaded that alternative manuals of translation can exist, incompatible with each other.[7]

Another problem with this interpretation is that, if the indeterminacy thesis is to function as part of a *reductio* of mentalistic conceptions of meaning, it would have to be interpreted as an epistemological thesis rather than an ontological thesis.

Thus, consider the case of a mentalist who accepts Quine's contention that there is no empirical (observable) fact to favor one rival manual over the other. Indeed, let us suppose that he even accepts the additional claim that there is no physical fact to favor one rival manual over the other. Such a mentalist may still maintain that there is a fact of the matter about which manual is correct. He may still maintain that there is some mental fact in virtue of which one manual rather than the other is correct. Such a mentalist may claim that the question of whether the native's term 'gavagai' is correctly translated by 'rabbit' rather than 'undetached rabbit part' depends upon what idea he (the native) associates with the term 'gavagai'. True, there may be nothing which could give us reason to favor one translation over another, nevertheless there is a fact of the matter about which translation is correct.

---

[7] "Mind and Verbal Dispositions," in *Mind and Language*, S. Guttenplan, ed. (New York: Oxford, 1975), pp. 82–95, here p. 90.

For the mentalist, the indeterminacy and inscrutability theses would have, at best, to be epistemological theses not ontological theses. Yet Quine himself explicitly rejects this reading. Thus, he says:

> The point is not that we cannot be sure whether the analytical hypothesis is right, but that there is not even . . . an objective matter to be right or wrong about.[8]

Thus, it seems Quine is faced with the following dilemma:

Either

(1) We approach the indeterminacy and inscrutability theses from a Quinean perspective, in which case there is no fact of the matter as to whether there is in fact any indeterminacy of translation or inscrutability of reference.

or

(2) We approach the indeterminacy and inscrutability theses from a mentalist's perspective, in which case translation and reference are at best empirically underdetermined, though not ontologically indeterminate or "inscrutable" (in the Quinean sense that "there is nothing there to scrute").

IX TOWARD A NEW INDETERMINACY THESIS

Clearly, one of the purposes of the indeterminacy thesis is to contrast translation theory with physical theory. For Quine, the former is both underdetermined and indeterminate, whereas the latter is at most underdetermined. For Quine, qua physical realist, there is a physical reality in virtue of which two empirically equivalent physical theories can have different truth values.[9] On the other hand, Quine does not believe that there is some semantic/intensional reality in virtue of which two empirically equivalent translation manuals can have different truth values.

---

[8] *Word and Object*, p. 73.

[9] Cf., however, "Things and Their Place in Theories," in Quine's *Theories and Things* (Cambridge: Harvard, 1981). In that place (cf p. 22) Quine seems to jettison his realism, claiming that a theory that accords with all observational data could not be false.

To put the point another way, Quine, qua physical realist, believes that truth is ultimately determined by the states and distributions of elementary particles. Two theories that fit the same states and distributions of elementary particles are, for Quine, not rival theories. Indeed, Quine even goes so far as to claim

> Two sentences agree in objective information, and so ***express the same proposition,*** when every cosmic distribution of particles that would make either sentence true would make the other true as well. Each distribution of elementary particles of specified kinds over total space-time may be called a possible world; and then two sentences ***mean the same proposition*** when they are true in all the same possible worlds.[10]

Yet, as Dagfinn Føllesdal[11] reports, Quine says of the rival translation manuals required for the indeterminacy thesis that they "fit the same states and distributions of elementary particles" (*ibid.*, p. 295). Indeed, Quine himself makes the point quite explicitly in his article, "Things and Their Places in Theories":

> When I say there is no fact of the matter, as regards, say, two rival manuals of translation, what I mean is that both manuals are compatible with all the same distributions of states and relations over elementary particles. In a word, they are physically equivalent (p. 23).

How can there be ***rival*** manuals that, by Quine's own lights, "express the same proposition" and "mean the same proposition"? Surely, if two manuals express the same proposition, they are not really rival manuals!

Since Quine is a fervent supporter of the old physicalist catch phrase, "no difference without a physical difference," he would, we expect, claim that any two manuals that are empirically equivalent and fit the same states and distributions of physical particles are not conflicting, hence not rival, manuals. Yet, in fact, his favored versions of the indeterminacy thesis entail that certain such manuals are conflicting/rival manuals!

In *Word and Object,* the following formulation of the indeterminacy occurs:

> The thesis is then this: manuals for translating one language into another can be set up in divergent ways, all compatible with the totality of speech

---

[10] *Philosophy of Logic* (Englewood Cliffs, NJ: Prentice-Hall, 1970), p. 4 [emphasis mine].

[11] "Indeterminacy of Translation and Underdetermination of the Theory of Nature," *Dialectica,* xxvii (1973)· 289–301.

> dispositions, yet incompatible with one another. In countless places they will diverge in giving, as their respective translations of a sentence of the one language, sentences of the other language which stand to each other in *no plausible sense of equivalence however loose* (p. 27, emphasis mine).

The inclusion of the italicized phrase makes this an especially strong version of the indeterminacy thesis. Indeed, it is so strong as to be trivially self-refuting. In particular, since such "rival manuals" are both "compatible with the totality of speech dispositions," they are obviously equivalent in some sense. Ordinarily, Quine would call this empirical equivalence. True, Quine here writes of the equivalence of sentences not of manuals. But note, the sentences of both manuals are equivalent as totals. That is, where *A* is translator 1's translation of *X* and *B* is translator 2's translation of *X*, *A* and *B* are equivalent in the sense that they, embedded in 1's manual, or, more properly, 1's total world theory, and *B* embedded in *B*'s total world theory have the same empirical import. To demand empirical equivalence at the level of a single isolated sentence would be to violate Quine's celebrated holistic strictures. The sense of empirical equivalence here is that which applies in the case of the German who says "Giovanni's Satz 'E un libro' bedeutet 'Es ist ein Buch' " and the Australian who says " 'Giovanni's sentence 'E un libro' means 'It's a book'." Their translations only count as empirically equivalent provided the utterers have appropriate background theories. For instance, suppose our Australian's background theory contains the claim 'Books are large marsupials whose primary means of locomotion is the hop'.

This little contretemps is not merely the result of an infelicitous statement of the indeterminacy thesis. Rather, it is indicative of the problem that is the center of this current work: Quine states the indeterminacy thesis as a claim about the existence of rival, that is, incompatible, empirically equivalent translation manuals, yet he cannot give any adequate gloss on the nature of the alleged rivalry/incompatibility. The manuals are not empirically rivals/incompatible, since they both fit all the observational evidence. They are not physically rivals/incompatible, since they are consistent with the same physically possible worlds. They are not formulations of conflicting theories, because they are empirically equivalent and intertranslatable. The only sense in which they are rivals is in the sense that, given the ordinary notion of meaning, they assign conflicting meanings to the same sentence. Yet this is the very notion of meaning Quine has been so trenchantly attacking, often using the indeterminacy thesis as the spearhead of attack!

Ironically, these considerations suggest the following formulations of the point behind the indeterminacy thesis:

(I.T.2) Any two empirically equivalent translation manuals that are apparently rival manuals are in fact merely different manuals.

and/or

(I.T.2) There cannot be two rival translation manuals, both of which fit an individual's speech.

How much justice do (I.T.1) and (I.T.2) do to Quine's views?

Consider that one of the things that typically separate physical realists from positivists and other antirealists is the question of whether having the right empirical implications is sufficient for the truth of (the statements of) a physical theory. Physical realists believe that it is what a theory states about physical phenomena, be they macroscopic or microscopic, and the nature of those phenomena, which determines the truth value of (the statements of) the theory. It is this which allows for the possibility of empirically equivalent physical theories having different truth values; despite their empirical equivalence, such theories may be making conflicting claims about the nature of physical reality, and hence may be rival theories.

Typically, antirealists claim that any two physical theories that have the same empirical implications are in fact identical theories and hence share the same truth value.

Analogously, we might say that one of the questions that typically separate mental realists from mental antirealists is the question of whether having the right empirical implication is sufficient for the truth of a mental theory (e.g., a theory about what the words of some individual's idiom mean). Mental realists will typically deny that having the right empirical implications is here sufficient for truth. By the same token, they will typically deny that any two mental theories that are empirically equivalent are in fact identical theories. On the other hand, mental antirealists will typically affirm both these claims.

Clearly, any consistent antirealist about intentional phenomena will be an advocate of (I.T.1) and (I.T.2) and Quine is notoriously an antirealist about intentional phenomena. This at least is some evidence for the claim that (I.T.1) and (I.T.2) do some justice to Quine's views.

KEN GEMES

Yale University

BOOK REVIEWS

Arnauld and the Cartesian Philosophy of Ideas. STEVEN M. NADLER. Manchester: University Press, 1989. ix, 195 p.

Steven M. Nadler tells us in his preface that this study "is intended to fill in an egregious gap in scholarship on the history of early modern philosophy." His careful analysis of Arnauld's *Des vraies et des fausses idées* (1683) goes a long way toward filling that gap. Arnauld's book on ideas is an important contribution to perception theory. His extended exchange with Malebranche should have received attention long ago by anyone interested in perception. We are indebted to Nadler for helping to reinstate Arnauld in the philosophical canon with this detailed and judicious analysis of Arnauld's text.

The focus of Nadler's attention is on the epistemic and ontological issues between Arnauld and Malebranche, issues which raised questions about the correct interpretation of Descartes on the nature of ideas. There were other disagreements between Arnauld and Malebranche over theological issues, some of which are indirectly related to the main topic of *Des vraies et des fausses idées.* Nadler helps the reader to understand the interrelations between the theology that Arnauld rejected in Malebranche and the theory of ideas (ch. I, §§1,2; Appendix). The connection lies in the role of God in the world, more particularly, his role in our acts of perceiving and our acts of moral choices. In both, Malebranche has God operating in accordance with general laws he established at the creation, laws which express his will. God's "volontés générales" use particular acts of men as the occasions for his causal intervention. The question was, can we choose and act without God's direct help? The general laws govern or express the relation between mind and body, but the question much debated in the eighteenth century was precisely what *is* that relation. For example, what is the difference between occasionalism and pre-established harmony? Nadler's interesting claim is that Malebranche differs from Arnauld (at least as Arnauld saw that difference) in making God's relation to and interaction with the world indirect, both for action and for perception. This is the difference between "Arnauld's direct realism and his theory of efficacious grace" *and* "Malebranche's *êtres représentatifs* and his account of God's *volontés générales*" (183).

Nadler aligns his interpretation of Arnauld's account of perception and ideas with some recent writers who have been arguing that Arnauld was a direct realist. That interpretation had been argued

0022-362X/91/8802/109–112

against A. O. Lovejoy in the 1920s by John Laird. The label is less important than what it is taken to designate. What Arnauld (and Locke) attacked in Malebranche was his concept of ideas as separately existing entities, as *beings* that God gives to us on appropriate occasions of sensory stimuli. The entities mediate between objects and perceiver. Their role was to render the physical world intelligible to us; they were not little simulacra of objects, but rather what we would call "concepts." For Malebranche, everything had to be either a substance or a quality. His ideas clearly fell on the side of substance, even though there are difficulties for his ontology in so extending the category of substance. Nevertheless, his ideas are entities.

In attacking Malebranche's representative beings, Arnauld was not rejecting the representative feature of ideas. He believed Malebranche was confused over the difference between *cognitive* and *physical* presence. Malebranche was led to his representative *beings* through his belief that direct, immediate presence of objects to cognizers was required. Obviously, physical objects cannot be present to the mind, since, as he reasoned, only immaterial objects can be present to an immaterial soul or mind. What Arnauld in effect did was to argue that the cognitive relation *is* able to cross ontological categories. Descartes's (and some scholastics) notion of *objective* being is understood by Arnauld as a distinctive kind of being, the *being known* by a perceiver of objects (see 147–165). Nadler suggests that we can understand this notion in terms of *intentionality* (165–178). Intentional objects, cognitive presence, do not by themselves by-pass representation. What they eliminate is the special entity, Malebranchian ideas, between perceiver and object. There is still, as it were, something between these two, namely, the cognition. Objects cannot be present to us cognitively without being known. The question is: How are we to analyze that cognitive relation, how does perceiving inform us about the world?

For Arnauld, it was the relation of representation that carried the cognitive content. His explication of that relation is not as helpful as we would like. Aside from a few examples of deriving more information about triangles from the idea of a three-sided figure, or understanding that a particular idea can be applied to many situations, or that the idea of God leads us to the idea of his existence, Arnauld does not give us much by way of analysis of that relation. He does make it clear that cognitive representation is unlike other kinds of representation (e.g., the way a painting represents a person): resemblance is not required. In fact, as Nadler points out, for Arnauld the cognitive relation of representation is *sui generis* and hence is "a primitive indefinable relation between the idea and the object it makes known" (174). Despite its indefinability, Nadler's analysis does help us understand what Arnauld was attempting to do, although I do not find Nadler's diagram on page 170 illuminating. Since to have an idea and to perceive are one and the same for Arnauld, what he was arguing was that representation does not require entities: the representation holds between the act of perceiving and the object. Representation of this sort does not involve indirectness, unless we believe that the very fact of knowing or perceiving, the change in objects from existing to being known, the move from physicality to cognitivity renders direct realism impossible. Such a belief is, I would say, misleading, if not false. Arnauld was one of the first philosophers in the modern period to show that, when ideas are no longer taken for special objects (the *tertium quid*), the indirectness traditionally associated with representative theories disappears. This important feature of Arnauld's analysis is clearly illustrated in Nadler's exposition of the difference between Malebranche's representative *beings* and Arnauld's representative *acts* (ch. IV).

Besides the term 'represent', there is another key term in the exchange between Arnauld and Malebranche, 'perceive' or even 'see'. When Malebranche speaks of 'seeing all things in the mind of God', he tends to use the French '*voir*', in contrast with '*regarder*', for 'looking at' or seeing in a sensory way. This difference of terminology is not always consistently followed, but it occurs often enough to take these words as designating different though related processes. Nadler suggests that, for Malebranche, what we see (*voir*) are those special intelligible objects in the mind of God. 'Perception' proper refers to that kind of 'seeing'. But we must be careful not to be misled into thinking that such 'seeing' is anything like the 'seeing' of trees, tables, etc., as we ordinarily use that term. '*Voir*' really is a form of *understanding*. Malebranche's special idea objects help us understand sensory objects as instances of, e.g., extension. They supply us with categories for making a kind of sense (geometrical sense) of the ordinary objects we see (*regarder*). As Nadler points out, Malebranche does want to say we perceive ordinary objects, but only by means of his special intelligible objects (75–6). The latter, however, do not sharpen our vision, they clarify our understanding of certain features (general, geometrical features) of the former.

In the end, then, we may have to ask: "What should a theory of perception do?" Or we may ask: "Was the interest in perception shown by writers such as Descartes, Arnauld, Malebranche, and Locke at all similar to our interests today?" However we answer these questions, I think we can say that an analysis of the term 'representation' as used by these writers is relevant to some of today's discussions among philosophers and cognitive psychologists. There are even many similarities of problems and vocabulary between present-day writers and the way-of-ideas tradition. What we sometimes forget is that there has been a history of perception and knowledge, a history which is often relevant and enlightening, though the older terminology and metaphysics may sound quaint or scholastic to our ears.

Nadler's book, a well-written, textually and historically trustworthy study, should be read by anyone interested in perception theory and problems of representation. The result might even be to lead readers back to Arnauld's important text. It is to be hoped that Nadler will continue his research in the history of philosophy: he has a sensitive ear for historical texts, he is skilled at clear exposition, his analyses show a firm philosophical hand.

JOHN W. YOLTON

Rutgers University

NOTES AND NEWS

Analysis is pleased to invite entries for its first Essay Competition. Papers may be on any topic broadly in keeping with the traditional concerns of the journal. An upper limit of 5000 words, including references and footnotes, will be strictly applied. Entries must be received no later than May 1. Authors should be aged 30 or under on that date (though prospective entrants who believe that there are grounds for relaxing the age rule in their case may consult the editor). The winning essays, with a selection of other entries, will be published together as the October 1991 issue of *Analysis*. Entries and inquiries may be sent to the editor of *Analysis*, Peter Smith, Phil. Dept., Univ. of Sheffield, Sheffield S10 2TN, England.

The *International Journal for Philosophy of Religion* is pleased to announce that Eugene Thomas Long has been appointed Editor-in-Chief, succeeding Bowman Clarke. Persons submitting articles for consideration should consult a recent issue of the journal for style and send two copies of the article and two copies of a one hundred-word abstract to Prof. Long, Byrnes International Center, Suite 308-B, Univ. of South Carolina, Columbia, SC 29208.

Arguably the best textbook of its kind.

Just published from HBJ!

UNDERSTANDING ARGUMENTS

An Introduction to Informal Logic

Fourth Edition

Robert J. Fogelin and Walter Sinnott-Armstrong
Both of *Dartmouth College*
Paperbound, 1991

This popular introduction to informal logic is thoroughly updated and revised to be clearer and livelier than ever in a new *Fourth Edition.* Presentation of speech act theory and conversational implication in Chapter 1 and elsewhere shows students how arguments are both practical and useful. Methods of necessary and sufficient conditions now offer a simpler, more intuitive exposition of inductive reasoning than was previously available using Mill's Methods. Introductions to legal, moral, and scientific reasoning reveal differences and similarities among these areas. More detail on deep analysis in this *Fourth Edition* helps students shape arguments for evaluation. "That's just like arguing" method allows students to test for validity without formal apparatus. New discussion of general conditionals helps students understand the limits of the propositional logic and avoid common fallacies. A modern approach to syllogisms now precedes the classical approach, and discussion of statistical syllogisms helps students to apply generalizations. Many new exercises and discussion questions appear in every chapter.

HBJ

Harcourt Brace Jovanovich, Inc.
College Sales Office
7555 Caldwell Avenue, Chicago, IL 60648
(708) 647-8822

New from The Free Press

“This is a very satisfying philosophical life... Ray Monk's biography is a contribution not only to our understanding of Wittgenstein as philosopher and as person, but of philosophy as finally confessional when it is truly great.”
—Arthur C. Danto,
The Times Literary Supplement

LUDWIG WITTGENSTEIN
The Duty of Genius
Ray Monk

This is the first full-scale one-volume biography of Ludwig Wittgenstein. Monk's compelling portrait reveals the fascinating connection between the private life of a tortured man who never fully came to terms with his gifts, and the genius who in the epoch-making works *Tractatus-Logico Philosophicus* and *Philosophical Investigations* radically changed the fundamental nature of philosophical thought in our time.

“Well written, intelligent and moving, this is a first-rate biography, a biography that stands as both a fine piece of scholarship and a vivid portrait of a man who wanted to be ‘morally perfect.’ Monk communicates the passion—there is no other word for it—that Wittgenstein felt for philosophy, and he also conveys the extreme loneliness of his vocation, spending years on work that he often felt no other human being comprehended.”
—The New York Times Book Review

1990 ISBN: 0-02-921670-2 $29.95

STRUCTURES OF SOCIAL LIFE
The Four Elementary Forms of Human Relations
Alan Page Fiske

Fiske develops an encompassing but elegantly simple new theory of the psychological bases of social life. human activities are structured in accordance with four fundamental models· communal sharing, authority ranking, equality matching, and market pricing—and people use different permutations of these models in every domain of social life in all cultures.

“Fiske's tetrad will be pleasing to the grand theorists of social science, for he has synthesized almost everything worth synthesizing in social theory into an appealing theory about four irreducible prototypes for constructing social relationships”
—Richard A. Shweder, University of Chicago

1990 ISBN: 0-02-910345-2 $29.95

***For VISA, MasterCard, or American Express orders,
call toll-free 1-800-323-7445 or 1-800-257-5755
between 9am-5:30pm Eastern Time.***

New from The Free Press

“Pletsch's great virtue is being able to write a scholarly biography that is, in both argument and manner, completely accessible to the intelligent layman.”
—Paul Robinson, Stanford University

YOUNG NIETZSCHE
Becoming A Genius
Carl Pletsch

The cult of genius, the idea of the great man born to create new forms in a passionate agony—personified by the likes of Rousseau, Beethoven, Goethe, and philosopher Friedrich Nietzsche—dominated the imagination of 19th century Europe. In a fascinating biographical inquiry into Nietzsche's early life, Pletsch demonstrates that contrary to the romantic paradigm, Nietzsche self-consciously set out to become a genius by learning the role from others. Pletsch argues that genius is a role that must be learned before it can be performed, and Nietzsche's unusually long and difficult attachments to Schopenhauer and Wagner and his ultimate liberation from their influence is a uniquely revealing study of this process. Both a biographical introduction to the life and work of the most complex philosopher of the modern era and an exploration of the nature of genius, this book makes for a valuable intellectual drama while shedding new light upon Nietzsche's thought and works.

April 1991 ISBN: 0-02-925041-2 $22.95

COSMOPOLIS
The Hidden Agenda of Modernity
Stephen Toulmin

Since the 17th century, the vision of Cosmopolis, a society as rationally ordered as nature itself, has dominated the modern era, dictating a deep faith in the inevitability of progress and the certainty of scientific method, and fueling extraordinary advances in all fields of human endeavor. However, the failure of this vision to encompass the full diversity of humanity and nature in the 20th century has created confusion and disillusion in its wake. Seeking to resolve this crisis, Toulmin offers a new interpretation of Western intellectual history that humanizes our conception of modernity and reconciles the precision of scientific theory with the reality of human experience.

“*Cosmopolis* is simply the best discussion of modernity and postmodernity available today...A wonderfully hopeful book.”
—Robert N. Bellah, co-author of *Habits of the Heart*

1989 ISBN: 0-02-932631-1 $22.95

Fp THE FREE PRESS
A Division of Macmillan Inc., 866 Third Avenue, New York, NY 10022

Oxford

Ethics With Aristotle
SARAH BROADIE,
Yale University
In this incisive study, Broadie gives an argued account of the main topics of Aristotle's ethics: *eudaemonia*, virtue, voluntary agency, practical reason, *akrasia*, pleasure, and the ethical status of *theoria*. Fresh light is shed on his comparison of practical wisdom with other kinds of knowledge, and a realistic account is developed of Aristotelian deliberation.
1990 480 pp. $55.00

Persons, Rights, and the Moral Community
LOREN E. LOMASKY,
University of Minnesota, Duluth
"An important event in the philosophical development of classical liberal theory.... Readable, entertaining, and far too full of moral truths to be confined to the artificial world of the academic." —*Reason* "An important work. Even those disagreeing with its central thesis will find it challenging." —*American Political Science Review*
1990 296 pp. paper $12.95

Aristotle's First Principles
TERENCE IRWIN,
Cornell University
"Undoubtedly a major work—scholarly, learned, detailed.... Something for Aristotelians to discuss for a long time." —*Times Higher Education Supplement.* Irwin here shows how Aristotle defends dialectic against the objection that it cannot justify a metaphysical realist's claims.
1990 720 pp. paper $24.95

Harmless Wrongdoing
JOEL FEINBERG, *University of Arizona, Tucson*
"In its scope, attention to principle and detail, careful argument, and respect for psychological and social realities, this last volume, as well as its predecessors, provides a first-rate analysis of which conduct the state may properly declare criminal and why. Feinberg's achievement, moreover, is independent of whether or not one agrees with his analysis or adopts his conclusions." —*Library Journal*
1990 384 pp. paper $14.95 cloth $34.00

Particles and Waves
PETER ACHINSTEIN,
John Hopkins University
"A terrific book. Rarely has such outstanding history of science been coupled with such subtle and penetrating philosophical analysis.... Out of that examination emerges Achinstein's carefully worked-out replacement analysis. Judged against the recent spate of very good historico-philosophical arguments, Achinstein's book stands out as making the most penetrating and profound philosophical contributions." —Frederick Suppe, *University of Maryland*
1991 360 pp. paper $24.95 cloth $49.95

Realism in Mathematics
PENELOPE MADDY,
University of California, Irvine
Offering a scrupulously fair treatment of both mathematical and philosophical concerns, Maddy here delineates and defends a novel version of mathematical realism. She answers traditional questions and poses a challenging new one, refocusing philosophical attention on the pressing foundational issues of contemporary mathematics.
1990 216 pp. $39.95

Prices are subject to change and apply only in the U.S.
To order, send check or money order to:
Humanities and Social Sciences Marketing Department
To order by phone using major credit cards please
call (212) 679-7300, ext. 7106

The Non-Reality of Free Will
RICHARD DOUBLE, *Edinboro University of Pennsylvania*
"An original contribution to debates about free will and responsibility, which discusses many recent developments in the field in a clear and engaging fashion and challenges every familiar position, often with novel arguments. It should provoke reactions from all sides of the debate and deserves to be widely read and discussed."—Robert H. Kane, *University of Texas at Austin*
1990 272 pp. $32.50

Hume's System
An Examination of the First Book of His Treatise
DAVID PEARS, *University of California, Los Angeles*
Focusing on the first book of the *Treatise*, Pears examines in detail its application to three difficult problems: causation, personal identity, and sense perception. Hume's solutions, Pears argues, are not theories that can be given a place in standard classification of philosophical theories, but rather depend upon a subtle form of naturalism not altogether unlike Wittgenstein's naturalism.
1991 224 pp. paper $12.95 cloth $35.00

Kant's Theory of Self-Consciousness
C. THOMAS POWELL, *Guilford College*
In this study, Powell shows that the structure of the Paralogisms provides an essential key to understanding both Kant's critique of "rational psychology" and his theory of self-consciousness. Powell's examination of Kant's theory of self-consciousness is also shown to have implications for contemporary discussions of the problem of other minds, functionalism, and the problem of indexical self-reference.
1990 280 pp. $56.00

Reading Nietzsche
Edited by ROBERT C. SOLOMON and KATHLEEN M. HIGGINS, *both of the University of Texas, Austin*
"By treating each book separately and focusing on the strengths, weaknesses, and coherence of each individual work, this collection of accessible essays may prove more useful to first-time readers of Nietzsche than anthologies which address Nietzsche thematically or topically.... This anthology is enthusiastically recommended."—*Choice*
1990 272 pp. paper $12.95 cloth $32.50

Demonstratives
Edited by PALLE YOURGRAU, *Brandeis University*
This collection brings together for the first time an important set of previously published papers which seek to provide the correct semantic account of thoughts involving demonstratives, as well as insight into the metaphysical issues that such accounts must speak to: for "I", the nature of the self, and for "now", the reality and significance of the present.
1990 288 pp. paper $12.95 cloth $39.95

New in paperback!
Aristotle's Two Systems
DANIEL W. GRAHAM, *Brigham Young University, Utah*
"An important book. Graham argues, in much greater detail than any previous scholar, that Aristotle's early logical treatises (the *Organon*) are incompatible, indeed, incommensurable with his later physical-metaphysical works.... *Aristotle's Two Systems* is one title with which every researcher in the field will have to deal. Highly recommended."
—*Choice*
1990 384 pp. paper $24.95

Oxford University Press
200 Madison Avenue • New York, NY 10016

Philosophy from Cambridge

Account Episodes

Peter Schönbach

Responsibility and accountability are the issues at the heart of *Account Episodes*. Schönbach defines an account as an explanation offered following a failure on the part of the person he refers to as "actor" to perform an action successfully.

European Monographs in Social Psychology Copublished with the Maison des Sciences de l'Homme

35017-4 Hardcover $49.50

The Theory and Scholarship of Talcott Parsons to 1951

A Critical Theory

Bruce C. Wearne

Drawing on unpublished papers, Wearne's timely study describes Parsons' religious background and his education, and traces the impact of German social theory on his development as a scholar and concludes with a thorough and fresh reading of the classic *The Social System*.

37003-5 Hardcover $39.50

Metaphors of Mind

Conceptions of the Nature of Intelligence

Robert J. Sternberg

Readers will gain a comprehensive understanding of the concept of intelligence and its continuing evolution from this text's perspectives of psychology, anthropology, computational science, sociology and philosophy.

35579-6 Hardcover $44.50
38633-0 Paper $17.95

Philosophical Papers

F.P. Ramsey

D.H. Mellor, Editor

A compilation of all previously published writings on philosophy and the foundations of mathematics from the greatest of the generation of Cambridge scholars that included G. E. Moore, Bertrand Russell, Ludwig Wittgenstein and Maynard Keynes

37480-4 Hardcover $49.50
37621-1 Paper $15.95

Rationality and Dynamic Choice

Foundational Explorations

Edward F. McClennen

"This is a seminal work which demands the serious attention of all those interested in rational choice."
—*David Gauthier*, University of Pittsburgh

36047-1 Hardcover $39.50

The Development of Newtonian Calculus in Britain, 1700-1810

Niccolò Guicciardini

"...provides a much needed reminder that mathematical ideas are transmitted through teaching and cultural influences as well as through research."—*Nature*

36466-3 Hardcover $54.50

Moral Philosophy from Montaigne to Kant: An Anthology

Volume I and Volume II

J.B. Schneewind, Editor

A major anthology of writings from the seventeenth and eighteenth centuries includes the works of Hobbes, Hume, Butler, Bentham and Kant, as well as those of a number of important but generally ignored writers.

V. 1 35361-0 Hardcover $49.50
35875-2 Paper $14.95
V. 2. 35362-9 Hardcover $49.50
35876-0 Paper $13.95

Now in paperback...

Augustine on Evil

G.R. Evans

"Dr. Evans' book can be recommended to anyone seeking a way into the Augustinian world in general...[it] fills a notable gap in the literature."
—*Religious Studies*

39743-X Paper $14.95

At bookstores or write.

CAMBRIDGE UNIVERSITY PRESS
40 West 20th Street, New York, NY 10011
Call toll-free 800-872-7423 outside NY State.
800-227-0247, NY State only. MasterCard/VISA accepted. Prices subject to change.

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NUMBER 3
MARCH 1991

page
113 *Epistemic Paternalism: Communication Control in Law and Society* Alvin I. Goldman
132 *Epistemology on Holiday* Mark Kaplan

BOOK REVIEWS
155 Jeremy Waldron: *The Right to Private Property* Alan Ryan
159 Richard W. Miller: *Fact and Method* Peter Kosso
163 Ronald N. Giere: *Explaining Science: A Cognitive Approach* Philip Kitcher

167 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1991 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER. Second class postage paid at New York, NY, and additional mailing offices Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NO. 3, MARCH 1991

EPISTEMIC PATERNALISM: COMMUNICATION CONTROL IN LAW AND SOCIETY*

A popular principle in epistemology and the philosophy of science is the requirement of total evidence (RTE). A weak version of this principle may be stated as follows:

> (W-RTE) A cognitive agent X should always fix his beliefs or subjective probabilities in accordance with the total evidence in his possession at the time [1]

This version says nothing about acquiring or collecting evidence, only about the use of evidence already in one's possession. A stronger version of RTE, however, addresses evidence gathering as well:

> (S-RTE) A cognitive agent X should collect and use all available evidence that can be collected and used (at negligible cost).[2]

This principle had best be understood in purely epistemic terms, rather than moral or legal terms. Some evidence collection might

---

* For valuable comments and advice, I am grateful to Holly Smith, Allen Buchanan, Henning Jensen, Joel Feinberg, Richard Lempert, and the members of the 1989 Dubrovnik conference on epistemology and philosophy of mind

[1] See Rudolf Carnap, *Logical Foundations of Probability* (Chicago: University Press, 1950), p. 211, and Carl G. Hempel, "Inductive Inconsistencies," in *Aspects of Scientific Explanation* (New York: Free Press, 1965), pp. 64–7.

[2] See I J. Good, "On the Principle of Total Evidence," in *Good Thinking* (Minneapolis: Minnesota UP, 1983); and Paul Horwich, *Probability and Evidence* (Cambridge: MIT, 1982), ch. 6. Good's argument for the strong version of RTE appeals to the criterion of maximizing expected utility, so it seems more pragmatic than purely epistemic. Horwich's argument, however, is purely epistemic. He shows that the *expected error* in one's probability judgment vis-à-vis some hypothesis H is minimized by the acquisition of any new evidence E (as long as $0 < \text{prob}(E) < 1$, and $\text{prob}(H/E) \neq \text{prob}(H/-E)$). I shall not discuss these arguments in detail because I am not so much interested in S-RTE as in the two-agent "control" version of RTE. In fact, I am not even primarily interested in the control version of RTE. I really use it as an expository heuristic to introduce the topic of epistemic paternalism, and the topic of communication control more generally.

0022-362X/91/8803/113–31

involve invasion of privacy or harmful experimentation on human subjects, which would be objectionable on moral and/or legal grounds. But if we abstract from these factors and restrict attention to epistemic considerations, the principle has initial intuitive appeal.

A plausible-seeming corollary, or extension, of S-RTE is a principle governing the practices of a second agent, *Y*, who is in a position to control the evidence made available to *X*. This interpersonal principle would say that *Y* should make available to *X* all evidence that is subject to his (*Y*'s) control. Of course, like S-RTE itself, the envisaged extension or corollary of S-RTE must be restricted to epistemic contexts or concerns. Thus, we might formulate the "control" version of RTE roughly as follows:

> (C-RTE) If agent *X* is going to make a doxastic decision concerning question *Q*, and agent *Y* has control over the evidence that is provided to *X*, then, from a purely epistemic point of view, *Y* should make available to *X* all of the evidence relevant to *Q* which is (at negligible cost) within *Y*'s control.

The restriction to the epistemic viewpoint is again important. In legal settings, for example, there are many nonepistemic reasons for refusing to provide relevant evidence to jurors. Available evidence may have been illegally obtained. Relevant evidence may be obtainable from the defendant, but the Fifth Amendment forbids his being compelled to testify against himself. Or the defendant may have testified elsewhere under a grant of limited immunity, providing that his testimony would not subsequently be used against him. In these cases, the judge (*Y*) is obliged not to provide the jurors (*X*) with all available evidence logically relevant to the question of guilt. These constraints, though, are not of an epistemic nature. Thus, the stated version of C-RTE may still be defensible. Indeed, anyone persuaded by John Stuart Mill's[3] famous thesis that, if we are interested in having truth prevail, we should allow all available arguments to be heard, should be attracted to C-RTE.

I shall argue, nonetheless, that C-RTE is unacceptable. More cautiously, I shall show that existing provisions and practices, both in the law and elsewhere in society, contravene C-RTE. Although I shall not defend each provision and practice in detail, many of them seem to be quite reasonable. This raises some interesting questions for a branch of epistemology that I have elsewhere called *social epistemics*.[4] The aim of this paper is to formulate these questions in a general way and to undertake a preliminary exploration of them.

---

[3] *On Liberty*, ch. 2.

[4] See *Epistemology and Cognition* (Cambridge: Harvard, 1986), pp. 1, 5–9, 136–8; "Foundations of Social Epistemics," *Synthese*, LXXIII, 1 (1987): 109–44;

I

The American legal system has a set of explicitly formulated rules of evidence for the federal courts. These and other trial-related procedures are substantially dedicated to the goal of getting the truth about the issues under litigation (in a criminal trial, the guilt or innocence of the accused). Since the aim of these rules is, accordingly, heavily or predominantly epistemic, they provide excellent examples for our scrutiny.

The Federal Rules of Evidence[5] resulted from proposals by an Advisory Committee appointed by the Supreme Court, which were passed into law (with some revisions) by Congress in 1975. Although the rules do not, in general, bind judges on the admission or exclusion of evidence, they provide guidelines that should be followed if decisions are not to be overturned. Rule 102 declares their purpose: "These rules shall be construed to secure fairness in administration, elimination of unjustifiable expense and delay, and promotion of growth and development of the law of evidence *to the end that the truth may be ascertained* and proceedings justly determined" (emphasis added). The truth goal is implicit in the declaration (Rule 602) that witnesses may only testify to a matter of which they have "personal" (especially perceptual) knowledge, and in the general exclusion of hearsay evidence (Rule 801). Such rules codify what McCormick calls a "most pervasive manifestation" of the common law, viz., insistence upon "the most reliable sources of information."[6] Reliance on reliable sources, clearly, is more likely to produce accurate (i.e., truthful) judgments. The admission of testimony by experts (Rule 702) also exemplifies a dedication to truth, since experts are people presumed to know relevant truths. The requirement of witnesses to make an oath or affirmation before testifying is still another device calculated to secure truthful evidence (Rule 603), as are statutory provisions of penalties for perjury. The entire procedure of cross-examination, which holds an exalted place in the Anglo-American trial system, is similarly rationalizable by reference to the truth goal. Since it is impossible to prevent false evidence from ever being introduced, at least there should be ample opportunity to contravert it, to reduce the likelihood of its being believed by the fact finder and thereby promoting a false verdict.

What is interesting in the present context is that the rules of

---

and "The Cognitive and Social Sides of Epistemology," in A. Fine and P. Machamer, eds., *PSA 1986*, II (East Lansing: Philosophy of Science Association, 1987), pp 295–311

[5] See *Federal Rules of Evidence for United States Courts and Magistrates* (St Paul: West, 1989).

[6] In E. Cleary, ed., *McCormick on Evidence*, 3rd ed. (St Paul: West, 1984), p. 23

evidence frequently foster the exclusion of evidence from the jurors, in conflict with principle C-RTE. On the matter of admission or exclusion of evidence, one governing provision is that *relevant* evidence is generally admissible, but *irrelevant* evidence is not admissible (Rule 402).[7] In the matter of hearsay evidence, the basis for exclusion ostensibly falls under a different principle, not irrelevance but unreliability or doubtful veracity. Still other rules promote exclusion even when the evidence is both relevant and presumptively true. Evidence about the character of an accused is not admissible for the purpose of proving that he acted according to character on the occasion in question (Rule 404). In particular, evidence of previous *crimes* by the accused is not admissible to help prove that he committed the present crime. Also, evidence that the defendant initially entered a guilty plea on the present charge, and then withdrew it, is inadmissible (Rule 410). Here we have examples of a particularly interesting kind, in which judges are allowed or enjoined to exclude presumptively true and relevant evidence. Can these be rationalized on epistemic grounds?

The general rationale is given in Rule 403, which states in part: "Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, ..." The phrase 'misleading the jury' apparently refers to the jury's being led into making an incorrect judgment, and the language of 'prejudice' and 'confusion' can also be readily interpreted this way. It appears, then, that the Federal Rules of Evidence provide an *epistemic* rationale for excluding various types of evidence from doxastic decision makers (jurors), contrary to what C-RTE enjoins.

II

A better understanding of the underlying rationale for the relevance rules can be gleaned from a theoretical reconstruction by a legal scholar, Richard Lempert.[8] Lempert offers several reconstructive suggestions, but I shall present just the one employing a Bayesian interpretation. The discussion assumes that the fact finder is a jury and the issue to be resolved is a defendant's guilt.

One form of Bayes' Theorem, suitable to the present case, is formulated as follows:

$$O(G/E) = \frac{P(E/G)}{P(E/\text{not-}G)} \cdot O(G)$$

---

[7] Relevant evidence is characterized as "evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence" (Rule 401).

[8] "Modeling Relevance," *Michigan Law Review*, LXXV (1977) 1021–57.

This formula describes the way that a new item of evidence (E) should influence a rational agent's odds (O) that a defendant is guilty (G).[9] It says that the posterior odds of the defendant's guilt is equal to (1) the probability that the evidence would obtain if the defendant is in fact guilty, (2) divided by the probability that that evidence would obtain if the defendant is in fact not guilty, (3) multiplied by the prior odds on the defendant's guilt. The ratio of items (1) and (2) is conventionally called the *likelihood ratio*. On Lempert's reconstruction, a necessary and sufficient condition for evidence to be logically *irrelevant* to the defendant's guilt is that the likelihood ratio be 1:1 (or close to 1:1). This holds when the evidence in question would be just as likely to arise if the defendant were not guilty as if he were guilty, i.e., if $P(E/G) = P(E/\text{not-}G)$. Evidence of this sort has no "probative" value, that is, warrants no change in the posterior odds of guilt. Such evidence is unhelpful and therefore should not be admitted. Logically *relevant* evidence is evidence whose likelihood ratio departs (substantially) from 1:1.

Why should courts declare even relevant evidence inadmissible? Lempert expresses this in terms of the danger that the fact finder will misestimate the probabilities that comprise the likelihood ratio, and hence assign excessive (or insufficient) weight to the evidence. Overestimating the numerator or underestimating the denominator makes the conclusion sought by the proponent of the evidence appear more probable than it actually is; underestimating the numerator or overestimating the denominator has the opposite result. For example, suppose that in an assault case it can be shown that the defendant is a heroin addict, and also that one out of 500 criminal assailants are heroin addicts, whereas of the people who never engage in criminal assault only one in 1,000 are heroin addicts. Then knowledge that the defendant is an addict should result in a doubling of the prior odds that the defendant was the assailant. Suppose further that the fact finder (mistakenly) thinks that the probability that a nonassailant would be a heroin addict was one in 10,000 rather than one in 1,000. This misestimation of the denominator by a factor of ten leads to a twentyfold increase in the odds of guilt rather than a twofold increase. It is precisely this sort of danger which concerns the courts. Rule 403, excluding character evidence, is partly justified on this ground, says Lempert. Jurors are likely to magnify the import of character traits, or past criminal record;[10] and similarly for withdrawn guilty pleas. When courts or the codified rules speak of the problem as one of juror "prejudice," or when they speak of "confusing" or "misleading" the jury, this should be understood, according to Lempert, in terms of the jury's being prone to such misestimations.[11]

---

[9] I do not mean to endorse the Bayesian analysis for all purposes, including legal purposes. I am merely expositing Lempert's reconstruction. The same morals could well be drawn without commitment to Bayesianism. For an alternative account of reasoning principles for the law, see L. Jonathan Cohen, *The Probable and the Provable* (New York: Oxford, 1977) (which I also do not mean to endorse).

III

Whether or not Lempert's Bayesian reconstruction is wholly faithful to the intent of the rules, it is apparent that the framers of the rules, and judges themselves, often wish to *protect* jurors in their search for truth. If, in the framers' opinion, jurors are likely to be misled by a certain category of evidence, they are sometimes prepared to require or allow such evidence to be kept from the jurors. This is an example of what I shall call *epistemic paternalism*. The general idea is that the indicated rules of evidence are designed to protect jurors from their own "folly," just as parents might keep dangerous toys or other articles away from children, or might not expose them to certain facts. I do not wish to quibble here over precise definitions of the term 'paternalism'. My usage, however, has much in common with H. L. A. Hart's characterization of paternalism as "the protection of people from themselves," and with Joel Feinberg's emphasis on the analogy to parental relations with children.[12] Jurors may have flaws in their background beliefs, or in their ability to draw apt conclusions from evidence. If so, the courts are prepared to protect them against these information-processing deficiencies in order to get truthful judgments on the issues at hand. Admittedly, in the judicial case, the prime objects of "protection" are plausibly the parties to the litigation, or perhaps society at large, not the jurors. So the present use of the term 'paternalism' may extend a bit beyond its standard usage. The indicated parties are protected, however, by getting jurors to make accurate judgments. Protection of the *jurors'* epistemic ends therefore assumes derivative importance.

---

[10] In defending the rule against character evidence, the Advisory Committee for the Federal Rules of Evidence quotes the California Law Revision Commission with approval: "Character evidence is of slight probative value and may be very prejudicial. It tends to distract the trier of fact from the main question of what actually happened on the particular occasion It subtly permits the trier of fact to reward the good man and to punish the bad man because of their respective characters despite what the evidence in the case shows actually happened" (*op. cit.*, p 31).

[11] In Horwich's terms (see fn. 2 above), we might express the court's fear of jurors being misled by E as an assessment that the result of their getting E would be an *increase* in the probability of their making an error rather than a reduction The crucial point, of course, is that the court substitutes its *own* expectation of juror error for that of the jurors themselves, who perhaps would welcome information about the accused's character, criminal record, or withdrawn guilty pleas as likely to reduce their probability of error.

[12] Hart, *Law, Liberty, and Morality* (New York. Vintage, 1966), p. 31; and Feinberg, *Harm to Self* (New York. Oxford, 1986), pp. 3–8 For other discussions of the definition of paternalism, see Rolf Sartorius, ed., *Paternalism* (Minneapolis. Minnesota UP, 1983); and Donald Van DeVeer, *Paternalistic Intervention* (Princeton: University Press, 1986), pp. 16–24.

In any event, I shall construe epistemic paternalism in a broad sense. I shall think of communication controllers as exercising epistemic paternalism whenever they interpose their own judgment rather than allow the audience to exercise theirs (all with an eye to the audience's epistemic prospects). Thus, the exclusion of evidence of doubtful veracity (e.g., hearsay evidence) also qualifies as epistemic paternalism. The courts apparently feel that jurors cannot be counted on to discount hearsay testimony adequately. So they substitute their own wisdom for that of the jurors. Similarly, when judges exclude evidence as irrelevant, they must use *their* assessment of whether the evidence affects the probabilities of the propositions in question (the jurors' assessments might have been different). When these categories of exclusion are added, it is clear that courts engage in a substantial amount of epistemic paternalism.

Is such paternalism really warranted? Are these rules *good* rules from an epistemic point of view? This is open to dispute. Lempert contends that evidence should not be excluded when the probative value is substantial, i.e., when the likelihood ratio deviates markedly from 1:1. The preferred solution, he says, is to provide the jury with the information needed to assess accurately the probative value of the offered evidence.

There are general problems with this solution, though. For one thing (as Lempert notes), there may not be hard data that indicate just how heavily a piece of evidence should be weighted. There may be no well-researched (base rate) "facts" to give to the jury. Nonetheless, people with judicial experience may have good reasons to suspect that juries would exaggerate the import of that type of evidence. Second, there is a question of whether statistical information, even if it were available, would psychologically displace or override jurors' prior prejudices.[13] Third, there is a question of whether *all* the relevant considerations that bear on the import of the evidence should be given to the jurors. Should they be presented with the Bayesian framework in terms of which the evidence may be interpreted? Should they be informed of theoretical disputes over the correctness of Bayesianism, or the potential for its misapplication? Much of this material would surely confuse most jurors. The dangers of introducing mathematical techniques into the fact-finding process have been emphasized by Laurence Tribe.[14]

---

[13] Studies by Lee Ross and colleagues give evidence of a "perseveration" phenomenon, in which previously acquired beliefs are not eradicated even by new evidence showing that the original basis of those beliefs was ill-founded. See Richard Nisbett and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs, NJ: Prentice-Hall, 1980), ch. 8. I have suggested more generally that beliefs can never be "erased," but at best only "overridden." This allows for the presence of *credal residues*. See *Epistemology and Cognition*, pp. 223–6.

I am inclined to think that *some* paternalism is appropriate in this arena, although I shall not take a firm stand on specific policies. What I want to do is identify the questions clearly, and put them in the framework of a wider set of questions. What we have here is a set of rules or practices whose adoption has an impact on the truth values of the doxastic decisions that cognizers make. One question is: For each such rule, how good is its impact from a *veritistic* point of view, that is, in terms of the likelihood of getting truth and avoiding error? Would alternative rules or practices have better veritistic properties?

This type of question falls under what I call *social epistemics.* Epistemics generally is my (partly "reforming") conception of epistemology, with three main divisions: primary individual epistemics, secondary individual epistemics, and social epistemics. All three divisions seek to make evaluations in terms of veritistic ends, i.e., in terms of the effect on people's getting truths and avoiding errors. The objects or targets of evaluation, however, differ from division to division. Primary epistemics studies the veritistic properties of basic psychological processes. Secondary epistemics assesses the veritistic properties of learnable problem-solving methods, such as mathematical proof techniques, carbon-dating procedures, and the like. Social epistemics studies the veritistic properties of social practices, or institutional rules that directly or indirectly govern communication and doxastic decision. Judicial rules of evidence clearly fall in this last category.[15] The focus of the present paper is the class of institutional rules or practices that have a flavor of epistemic paternalism, e.g., those which weed out some possible communications that might be directed to a cognitive agent, by appeal to that agent's own veritistic ends. Are such rules and practices defensible, and, if so, which ones?

---

[14] "Trial by Mathematics: Precision and Ritual in the Legal Process," *Harvard Law Review*, LXXXIV, 6 (1971): 1329–93.

[15] For further discussion of the divisions of epistemics, and the rationale for them, see "Foundations of Social Epistemics" and "The Cognitive and Social Sides of Epistemology." For another treatment of judicial rules of evidence in the context of social epistemics, see Keith Burgess-Jackson, "An Epistemic Approach to Legal Relevance," *St. Mary's Law Journal*, XVIII, 2 (1986): 463–80. I thank Burgess-Jackson for helping draw my attention to relevance rules

Although social epistemics could content itself with the separate evaluation of individual practices, it seems more advisable to address whole classes of practices. What one wants is a general assessment of paternalistic control practices, first in terms of purely epistemic considerations and second on more inclusive grounds. Before turning to general principles, however, let us examine some additional cases of paternalistic practices.

IV

Start with curriculum selection in education, especially in primary and secondary schools. School personnel at various levels—boards of education, principals, and teachers—select curricular materials in the form of textbooks, course syllabi, and so forth. (When teachers do it, they assume something of a dual role: both controller and speaker.) In the nature of this business, some points of view and supporting argumentation are left out. Students are not exposed to all possible ideas on a given subject. Is this objectionable? Set aside highly publicized examples of book banning based on obscenity, profanity, or offensiveness. Think instead of simply ignoring opinions that have (or once had) their exponents, but are regarded by current authorities as palpably false or indefensible. Mathematics classes do not present alternative (crackpot) mathematics. Science classes do not (often) present the flat-earth viewpoint, Ptolemaic astronomy, or astrology. Schools rarely if ever invite Jeane Dixon or her ilk to give guest lectures, or recount as serious alternatives the theories of Velikovsky. Classes in health education do not give "equal time" to drug pushers to defend the safety of drug use, or to quacks to present and defend their cures. These omissions probably have veritistically good consequences. Of course, it might also be appropriate to have classes sample and compare "bad" and "bogus" science along with the good.[16] But this kind of course might be advantageous only for high-school students, not for younger pupils, who would only be confused.

The most public controversy surrounds the question of "equal time" for creationism in biology classes. Here, too, paternalism seems to be warranted on epistemic grounds. Experts on science should be allowed to decide that creationism is not a scientifically viable or serious contender, and hence should not be taught in the classroom.[17] What about the teaching of creationism in nonscience classes? Is it legitimate for a local community to require this in their schools? Here extra-epistemic issues intrude. Creationism is presumably a religious doctrine, so that constitutional issues arise about its being taught in public schools. What about the putative right of parents to control their children's education? If such an alleged right is supposed to rest on purely epistemic considerations, it would be very hard to defend. If it is based on nonepistemic considerations of an ethical or sociopolitical nature, it goes beyond the main focus of our discussion, although I shall briefly return to it in section VII.[18]

---

[16] The phraseology is borrowed from the title of Martin Gardner's book, *Science: Good, Bad, and Bogus* (Buffalo: Prometheus, 1981).

[17] This is not to dispute Paul Feyerabend's point [*Science in a Free Society* (London: Verso, 1978)] that the expertise of science and scientists must ultimately be judged by laymen. The thesis that laymen can assess the credentials of putative experts is discussed below in sect. VIII.

Turn next to the sphere of commercial advertising, where the Federal Trade Commission has authority to take actions against false or deceptive advertising. This exemplifies epistemic paternalism because it prevents some potential messages from being conveyed to an audience, and it does so with veritistic ends in mind: to keep them from believing untruths about commercial products. The FTC has four types of remedies for inhibiting false and deceptive advertisements, or even requiring truthful disclosures that the "speakers" would not independently make.[19] First, it can issue *cease and desist* orders. Second, it can seek to obtain *affirmative disclosure,* forcing a manufacturer to disclose by mark or label material facts concerning the merchandise. The most familiar such example is the requirement of warning labels on cigarette packages (although this is under the jurisdiction of the Food and Drug Administration, rather than the FTC). A third remedy is *corrective advertising*. When consumer misconceptions about a product have resulted from past advertising, the FTC has sometimes forced the manufacturer to devote 25% of its advertising budget for one year to corrections of past inaccuracies. In this arena, then, communication control sometimes takes a stronger form than the one considered thus far: not excluding messages, but mandating messages of a specified kind or content. A fourth remedy is the FTC's advertisement *substantiation* program. This forbids manufacturers to make certain advertising claims unless they have done competent scientific tests that substantiate their claims. This piece of epistemic paternalism has been rationalized by appeal to the relatively poor opportunity of consumers to obtain accurate information themselves.[20]

---

[18] On issues of educational philosophy and the control of educational policy, see Amy Gutmann, *Democratic Education* (Princeton: University Press, 1987)

[19] On these topics see Donald Gillmor and Jerome Barron, *Mass Communication Law,* 4th ed. (St. Paul: West, 1984), ch. 8; and Daniel Rohrer, *Mass Media, Free Speech, and Advertising* (Dubuque, IA: Kendall/Hunt, 1979), ch 3

[20] FTC Chairman Miles Kirkpatrick wrote: "The manufacturer has the ability, the knowhow, the equipment, the time and the resources to undertake such information by testing or otherwise—the consumer usually does not" (quoted in Gillmor and Barron, pp 618–9).

Another class of examples involves television and radio news. Generally speaking, American network news broadcasts offer a relatively limited variety of interpretations of each news event; nor do they detail the evidence for each of the possible interpretations. This is partly due, no doubt, to severe time constraints. But there is also the deliberate attempt to simplify, to make the news understandable and digestible to a large audience. Certainly this must be included under the rubric of epistemic paternalism. There is room for debate over how acceptable this practice is, and doubtless different programs and newscasters deserve different ratings on this dimension. But it is hard to deny the need for some degree of simplification, especially on technical topics. If this is at all legitimate, some degree of epistemic paternalism again seems inevitable and unobjectionable.

Simplification involves the omission of some truths. So paternalism in this form reduces the number of truths an audience has an opportunity to acquire. Nonetheless, the trade off of error avoidance, or confusion avoidance, may compensate for this loss (depending partly on how different segments of the audience are affected).

A further element in the trade off is that simplification can increase the audience size. This may mean that more true beliefs (i.e., belief tokens) are acquired through the simplified set of messages than would otherwise be the case. Although members of the media may care about this chiefly because of Nielsen ratings, and their consequent attractiveness to advertisers, the number of people acquiring true beliefs is a genuinely epistemic value, which I elsewhere call *fecundity*.[21]

The epistemic consequences of an institutional policy are often difficult to anticipate. The Federal Communication Commission's "Fairness Doctrine" is a good case in point. Since holders of its broadcast licenses are supposed to operate in the public interest, the FCC has required licensees to devote a reasonable amount of time to issues of public importance. Moreover, if a station presents one side of a controversial issue of public importance, it must afford reasonable opportunity for presentation of contrasting views. This policy was intended to promote veritistic outcomes by exposing the public to a wide diversity of opinion. It has been argued, however, that its actual result is opposite to its intent. The threat of incurring FCC suits has allegedly discouraged stations from airing controversial material. This "chilling" effect has tended to guarantee blandness, and to retard rather than enhance a free marketplace of ideas. This is a good example of how a policy's epistemic consequences (historical or prospective) can be a complex empirical question. This is why empirical sciences have a role to play in social epistemics.

---

[21] See "Foundations of Social Epistemics" and "The Cognitive and Social Sides of Epistemology."

V

Let us now inquire into the general circumstances in which epistemic paternalism, and other communication control policies, have good or bad epistemic consequences, i.e., good *veritistic* outcomes. A number of variables are relevant, especially: (1) the characteristics of the *controller* (or "gatekeeper"), (2) the characteristics of the *speakers* who wish to send messages via the communication channel, (3) the controller's *criterion of selection* among speakers or messages, (4) the characteristics of the *audience,* and (5) the availability of *alternate channels* that address the same topic.

Epistemic outcomes obviously depend heavily on the (epistemically) good judgment of the message controller. If the controller accurately distinguishes true and false claims, and/or true and false pieces of evidence for claims (and their relevance), then the choice of included and excluded messages may well promote truth acquisition on the part of the audience. This brings us immediately to variable (3): the criterion of selection. Does the controller select among candidate messages by reference to (A) their specific content, or (B) the characteristics of the speaker (or a combination of the two)? If the messages directly attempt to answer primary questions in the domain, and the controller himself has expertise on these questions, then a selection based on message content may positively contribute to veritistic outcomes. The controller may not purport to know answers to the primary questions in the domain, however. He may only select messages on the basis of prospective speakers' "credentials." Was the speaker an eyewitness to the putative events he wants to describe, or does he have some other access to, or authority about, the facts in question? If the controller is to make veritistically good selections, he must at least have expertise on the question of speaker credentials. (We might call this *secondary,* as opposed to *primary,* expertise.) This is what judges or rule framers claim to have in the field of legal evidence.

Selection criteria do not always involve purely veritistic considerations. A network may shy away from certain political interpretations because of prospective sponsor or audience disapproval. Since this is not a truth-oriented selection rationale, it raises doubts about the epistemic optimality of the control process.

Speaker characteristics are equally critical to (audience) veritistic outcomes. We have already mentioned the "access" or authority of speakers on the matters whereof they speak. Of almost comparable importance are the speakers' motivational properties. Do prospective speakers have an incentive for accuracy, or do they have incentives for deception and misrepresentation, as in the case of manufacturers, partisan witnesses, and perhaps even research scientists? If there are incentives for deception, does the *controller* know or suspect them, and does he use them in his selection policy?

Turning to audience characteristics, how informed, sophisticated, rational, etc., are the audience members who are likely to receive the channel's messages? How good are they at weighing competing views or complex chains of evidence? How good are they at assessing the reliability of various speakers or sources of evidence? Clearly, all these traits are crucial to veritistic outcomes. Furthermore, to the extent that the controller takes account of audience characteristics in selecting messages, the controller's knowledge (or true belief) about their characteristics is critical. If the controller has an accurate assessment of the audience's cognitive limits or shortcomings, a paternalistic choice of messages to include (or to *mandate*) may avert potential audience error. Or his policy of selection may avert potential *confusion*, which can lead to doxastic indecision or simple incomprehension of the truth (either of which is a comparatively bad veritistic outcome).

Finally, the availability of alternate communication channels can influence veritistic outcomes. If one channel deploys restrictive selection policies, that does not necessarily mean that the audience is confined to those messages. There may be other potential sources of information.

These are some of the salient factors which bear on the epistemic success of various alternative institutions of communication control. They are the sorts of factors to which social epistemics should appeal in evaluating existing or prospective institutions. As the discussion indicates, institutions or policies of these sorts cannot be rated in the *abstract*, apart from the properties of the individuals who occupy, or are likely to occupy, the different institutional roles. The same institution or policy might work well if the controller has considerable expertise, but poorly if he does not. In any case, when the cited variables or parameters take on appropriate values, epistemic paternalism will be justified. My previous examples are, quite plausibly, instances of this sort. So C-RTE must be rejected as a universal principle.

VI

I have been equating epistemically valuable outcomes with true belief and error avoidance. Are there additional epistemic values that this ignores? What about Mill's value of a "clearer perception and livelier impression of truth," produced by its collision with error (*op. cit.*, ch. 2)? And what about the skills to be learned from the process of wrestling with competing doctrines? I suspect that the livelier impression of truth, of which Mill speaks, can be cashed out in terms of a grasp of a larger number of related truths. One learns not simply that answer A_1 to question Q is correct, but that certain other answers, A_2, A_3, etc., have been offered to Q, what the arguments for and against these alternative answers are, and why these alternatives are mistaken. Learning these things doubtless has epistemic value, but it is not an additional ***kind*** of value. Similarly, the value of argumentative skills can be cashed out in terms of their instrumental value for discovering further truths on one's own, when no instructor is available.

Thomas Scanlon[22] has expressed doubts about epistemic protectionism by appeal to the value of autonomy, which consists in a person seeing himself as sovereign in deciding what to believe and in weighing reasons for action. Is autonomy an epistemic value of which a person is deprived by epistemic paternalism? If sovereignty is a matter of reserving the final doxastic choice to oneself, as some of Scanlon's passages suggest, this is not compromised by epistemic paternalism. The juror who did not hear certain excluded evidence still has to decide whether to believe what was.presented. Perhaps the point, however, is that, whenever a cognitive agent is deprived of some doxastic alternative, or some evidence relevant to that alternative, some degree of sovereignty is reduced. This may be granted, but does this reduction constitute a diminution in ***epistemic*** value (especially ***intrinsic*** epistemic value)? I doubt it, though the matter is not wholly clear. An analogous case involves "practical" value, and the range of alternatives actually scanned by a deliberating agent. Suppose agent *S* wishes to select a plan of action designed to achieve goal *G*. If an assistant fails to call *S*'s attention to some unnoticed alternative plan which is in fact worthless (it would not achieve *G* at all, or would be much too costly), *S*'s planning sovereignty has been reduced; he loses an opportunity to deliberate about the unpresented alternative. But is this a loss in "practical" value? Again, I am doubtful.

Even if we concede a measure of (intrinsic) epistemic value to autonomy, a value compromised in epistemic paternalism, it is questionable whether the loss always outweighs the gains. As John Hardwig[23] has stressed, we live in an epistemically complex world, where each of us cannot reasonably hope to assess all evidence for all theses personally. We often have to depend on the authority of others.

---

[22] "A Theory of Freedom of Expression," *Philosophy and Public Affairs*, I, 2 (1972): 204–26.

[23] "Epistemic Dependence," this JOURNAL, LXXXII, 7 (July 1985): 335–49.

Given this situation, it seems likely that epistemic paternalism will frequently be necessary, and sometimes epistemically desirable.

VII

The burden of the preceding sections is that epistemic paternalism can sometimes be warranted on *epistemic* grounds. Extra-epistemic considerations, however, should unquestionably enter into overall assessments of communication control policy. What are these other considerations?

Let me briefly suggest five additional factors that ought to be considered. First is the *practical significance* factor. Some communications induce audience beliefs that are likely to produce immediate actions with serious outcomes, including harms either to themselves or to other agents. (This is similar to the "clear and present danger" idea.) Direct and serious outcomes are salient in a trial, where juror beliefs dictate court sanctions that directly affect the defendant's welfare (and the welfare, one might say, of society).[24] Similarly, consumer beliefs about commercial products standardly lead to consumption choices, which may directly affect health or safety. These are the cases where epistemic paternalism seems most appropriate. Consumers' welfare, for example, seems to outweigh any presumed right of manufacturers to send any message they like.[25]

A second factor is the *power of the controlling agent*. Epistemic paternalism on the part of isolated individuals is quite a different matter from paternalism exercised by the state, or any other powerful organ of society. There are historical reasons for being very cautious about state control of information.

Closely related to this is the third factor: the *scope* of control. The greater the breadth, the greater the dangers of communicational restriction. This was already mentioned in section V, but is worth repeating here. The exclusion of certain viewpoints from school curricula seems more defensible when it is recalled that students have other communication channels available. They can be taught things (e.g., religion) outside of school. In the case of broadcast news, the public is free to select among numerous stations, and to read newspapers, books, and magazines as well. Thus, the extent of control of any single news outlet is *partial* (and small). The chief objection to state censorship, by contrast, is that it is *total*. This is a very different kettle of fish. My condoning of local epistemic paternalisms should not be taken as an endorsement of global state censorship, especially on topics like news, politics, and public affairs.

---

[24] Of course, where the defendant is in fact guilty, his welfare is not served by finding the truth. Presumably, however, society's utility is increased by accurate verdicts.

[25] There is ongoing debate, however, over the extent to which commercial speech is protected under First Amendment guarantees of freedom of speech. Not everyone agrees that regulation of advertising is legally defensible, or a good idea. See, for example, R. H. Coase, "Advertising and Free Speech," *Journal of Legal Studies*, VI, 1 (1977): 1–34.

This brings me to a fourth factor: the *topic* of the messages being controlled. It has been contended—quite plausibly, in my opinion—that an unregulated marketplace of ideas seems essential for political speech but not commercial speech. Thus, policies should differ for different subjects. I mention this here under nonepistemic factors; but actually it *may* be explainable by reference to epistemic ones. Perhaps we feel comfortable with a mathematics or science authority omitting nonstandard treatments from the curriculum because we acknowledge their expertise in their subjects. By contrast, in matters of (normative) politics and religion, we have no analogous conviction—as a society, at any rate—that there are any experts, and certainly no societal consensus on who the experts are (if any).

A fifth factor (or group of factors) concerns the putative *rights* of concerned agents: a speaker's right to freedom of speech, a communication channel owner's right to run his own station or publication, a parent's right or a community's right to predispose their children toward certain values or heritages, and so on. As indicated, some of these (putative) rights may be rationalized by epistemic factors, but others may be independent. In each case, moreover, the right may not be unconditional, but must be balanced by other rights or interests.[26]

How these nonepistemic factors should be balanced along with the (purely) epistemic ones is a complex matter, and I have no formula to offer. The variability of situations along these dimensions, however, makes it reasonable to expect epistemic paternalism to be appropriate (all things considered) in some contexts though inappropriate in many others.

VIII

Since expertise is a prime epistemic factor in the defense of epistemic paternalism, more should be said about it. To justify any particular instance of such paternalism, involving a particular controller, we must have grounds for taking that agent to be an expert. But how, it may be asked, are people supposed to *tell* who (if anyone) has expertise? In particular, if (primary) expertise is defined in terms of believing domain-relevant truths, how can people lacking expertise —novices—identify an expert? Furthermore, when different people hold conflicting views and each claims to be an expert, how can the issue be resolved to their mutual satisfaction?

---

[26] Scanlon has emphasized that issues in the freedom of expression hinge on the interests of three parties: participants (speakers), audiences, and bystanders. See "Freedom of Expression and Categories of Expression," reprinted in David Copp and Susan Wendell, eds , *Pornography and Censorship* (Buffalo: Prometheus, 1983). In cases of communication control, however, the interested parties are even more numerous. And in the case of children, some sort of special status (not merely "bystander") seems appropriate for parents and community

The problem of *expert identifiability* involves large epistemological issues that can only be addressed here partially, not fully. First, let us define an expert as someone who either (1) *has* true answers to core questions in the domain (i.e., believes or assigns high probability to these answers), or (2) has the *capacity* (usually through the possession of learned methods) to acquire true answers to core questions when they arise. Degree of expertise, then, is primarily a function of the candidate's question-answering *power,* i.e., how often he can form a belief in a correct answer as opposed to having either no opinion or an incorrect opinion.[27] Expertise can also be understood in either a *comparative* or an *absolute* sense. Someone is comparatively expert if his question-answering power ranks high compared with others; absolutely expert if his power ranks high in absolute terms.

The crucial requirement for expert identifiability is that, once candidate experts have produced their answers to target questions, there are "truth-revealing" situations in which novices and rivals can recognize which answers (if any) are correct. With enough such cases, the genuine experts can be identified. In fact, there are many domains in which such truth-revealing situations exist. Let me catalogue four homely domains of this sort: (1) *prediction* domains, (2) *factual-record* domains, (3) *repair* domains, and (4) *design* domains.

Prediction domains are ones in which the core questions are questions about future events, e.g., the weather, the economy, the stock market, or an upcoming election. When the events actually occur, the correct answers to the questions are revealed.

Factual-record domains can be illustrated with baseball and opera. An expert in these fields can correctly answer (from memory) questions about baseball or opera facts, e.g., who stole the most bases in 1973, which arias are from which operas, and so forth. Here there are undisputed records to which novices and rivals can appeal to verify correctness.

---

[27] The epistemic importance of question-answering power is emphasized in my *Epistemology and Cognition,* ch. 6. A further dimension in the definition of expertise should probably be added: the ability to recognize *wrong* answers as wrong (or assign them a low probability). Even if nobody has a correct answer to question *Q,* someone who knows (or truly believes) that each of the available answers is wrong is better off, epistemically, than someone who mistakenly believes (or assigns high probability to) wrong answers.

System-repair cases are more interesting. Whether a system (e.g., an air conditioner, an automobile, or a human body) is functioning properly or malfunctioning is frequently an uncontested, novice-detectable affair. Suppose system *S* is malfunctioning, so the question arises: "What techniques, steps, operations, or therapies would be followed by *S* returning to proper functioning?" A novice will typically have no answer at all. An expert, by contrast, can often answer correctly. The expert electrician, mechanic, or physician commonly proceeds as follows. First, he tries to find a system component that is causing the malfunction, using gauges, instruments, or other diagnostic techniques. Second, having diagnosed the cause, he decides what part replacements, alterations, adjustments, or medications would correct the malfunction. He thereby arrives at an answer to the target question. Can the truth of his answer be checked? Often it can. If and when the prescribed measures are executed, it will be novice-detectable whether the system resumes normal operation.

The fourth category of domains—design domains—is a fairly heterogeneous lot. Here the core question is: "What specific design or performance features would produce a certain desired effect, or meet certain standards or specifications?" This is the sort of question that sundry artists, designers, product engineers, or even mathematicians might be asked. The engineer might be asked for a blueprint of a dam that will meet certain specifications. A mathematician might be asked for a sequence of steps that would comprise a proof of a given theorem (according to understood constraints on the nature of proof).[28] In each of these cases, novices could not answer such questions correctly very often; but an expert could. (At least he could indicate correctly which answers are erroneous.) Is the correctness of an answer novice-detectable? Often the answer is 'yes'. Unable to produce a proof himself, the novice mathematician may nonetheless be able to verify whether someone else's proposal constitutes a proof. Whereas it may be difficult to design a theater set or choreograph a sequence that produces a desired effect, it may be palpable and uncontroversial that a certain design or sequence *does* (or does not) achieve that effect.

The moral is that, in many familiar cases, it is relatively easy, even for novices or rivals, to identify expertise. Needless to say, not all cases are straightforward. There is no necessary 'paradox', however, or impossibility in principle, of expert identification. There is no theoretical objection, then, to invoking the notion of expertise, construed veritistically, in the evaluation of communication practices.

---

[28] Even certain questions of theoretical science might be construed in this fashion, namely, "What is a theory in domain *D* that meets all the known empirical and conceptual constraints?"

IX

Let me return to the theme of epistemic paternalism, first illustrated with the rejection of C-RTE. To the extent that traditional epistemology has been attentive to the existence of multiple agents, it has generally assumed an idealized setting, in which all agents have the same cognitive resources, skills, and opportunities, where there are no time constraints, and so forth. In that sort of setting, a communication control principle like C-RTE might make sense. But in settings marked by different levels of expertise, by different opportunities for information gathering, by different levels of cognitive maturity and training, and by severe time constraints, idealized principles of communication do not plausibly apply. A social epistemology for the real world needs to take these constraints into account.[29]

The very question of communication control policy is relatively neglected in philosophy,[30] perhaps because the topic has been so thoroughly dominated by the Millian dedication to a free market of ideas. Interestingly, though, organized science and scholarship are very far from laissez-faire marketplaces. On the contrary, professional journals rely heavily on (putatively) expert referees and editors to weed out inferior contributions. Only offerings that are judged methodologically sound, well-informed, and possibly in the direction of truth are accepted for publication (or, in a related arena, given research funding). Thus, even in the purely intellectual sphere, laissez faire is not the de facto policy. More precisely, there is a complex layering of institutions. In free societies, governments pursue the laissez-faire policy of allowing private information channels to proliferate. But these information channels may themselves be highly restrictive in the messages they transmit. It could be argued that successful pursuit of epistemic ends depends not only on "deregulation" at the highest level, but on wise regulation at lower levels. This problem needs a more subtle and systematic examination than it has hitherto received. The present paper is intended to help motivate such an examination.

ALVIN I. GOLDMAN

University of Arizona

---

[29] I stress a similar theme of psychological realism in *Epistemology and Cognition*. Issues raised for moral theory by cognitive limits are stressed by Holly M. Smith in "Making Moral Decisions," *Nous*, XXI (1988): 89–108, and "Two-Tier Moral Codes," *Social Philosophy and Policy*, VII, 1 (1989). 112–32.

[30] One author who stresses the role of communication in a theory of rationality, however, is Jürgen Habermas, *The Theory of Communicative Action, vol. 1, Reason and the Rationalization of Society*, Thomas McCarthy, trans. (Boston: Beacon, 1981).

EPISTEMOLOGY ON HOLIDAY*

SOME years ago, an acquaintance complained to me about her first—and only—philosophy course. Her instructor had devoted a substantial portion of the course to the question of what—if anything—justified his students' belief that there was a lectern at the front of the classroom. To my acquaintance, the entire exercise had seemed a sham. It was obvious to her that there was a lectern in the front of the room and it seemed, as far as she could tell, that it was equally obvious to everyone else in the classroom, the instructor included. And it seemed to her intellectually dishonest of the instructor, and those students whose interest he had managed to engage, to pretend to throw into question the propriety of a belief when, in fact, the truth of that belief was entirely evident to them.

I had heard the complaint before—it was a complaint I had received from some of the students to whom I had taught epistemology. And it was a complaint to which I had a ready answer. The answer was Roderick Chisholm's.

I

In "Theory of Knowledge in America,"[1] and the subsequent two editions of his monograph *Theory of Knowledge*,[2] Chisholm contends that those who put forward such complaints are guilty of having misunderstood the question that is being asked them. Understood as a challenge to the propriety of the belief, the question 'What justification do you have for your belief that there is a lectern at the front of the classroom?' is indeed misguided. But, Chisholm continues, that is not how the question should be understood. We choose to focus upon the belief that there is a lectern at the front of the classroom precisely because we are antecedently convinced that we *are* justified in believing it is true. Our question about that belief, far from meaning to cast doubt upon its propriety, actually presupposes that our belief is justified. The question is simply intended to elicit the nature of that justification.

---

* I would like to thank Richard Foley, Miriam Solomon, William Wainwright, Joan Weiner, and the More or Less Exact Philosophy Discussion Group at the University of Cincinnati for helpful criticisms of earlier versions of this paper.

[1] Repr. in R. M. Chisholm, *The Foundations of Knowing* (Minneapolis: Minnesota UP, 1982), pp. 109–93, esp. pp. 130–1; orig in *Philosophy, The Princeton Studies: Humanistic Scholarship in America* (Englewood Cliffs, NJ: Prentice-Hall, 1964).

[2] Englewood Cliffs, NJ: Prentice-Hall, 1966, pp 24–5; in 2nd ed. (1977), pp. 16–8.

0022-362X/91/8803/132–54

Chisholm is quite explicit—and consistent over the course of the eleven years or so between the first and third of these writings—about how he conceives of this process of elicitation, what he calls the process of "investigating the theory of evidence from a philosophical—or Socratic—point of view."[3] I quote from the second edition of *Theory of Knowledge:*

> In order to formulate, or make explicit, our rules of evidence, we will do well to proceed as we do in logic, when formulating the rules of inference, or in moral philosophy, when formulating the rules of action. We suppose that we have at our disposal certain instances which the rules should countenance or permit and other instances that the rules should reject or forbid; and we suppose that by investigating these instances we can formulate criteria which any instance must satisfy if it is to be accepted or permitted, as well as criteria which any instance must satisfy if it is to be rejected or forbidden. To obtain the instances we need if we are to formulate rules of evidence, we may proceed in the following way.
>
> We consider things that we know to be true, or think that we know to be true, or certain things which, upon reflection, we would be willing to call evident. With respect to each of these, we then try to formulate a reasonable answer to the question, "What justification do you have for thinking you know this thing to be true?" or "What justification do you have for counting this thing as something that is evident?" In beginning with what we think we know to be true, or with what, after reflection, we would be willing to count as being evident, we are assuming that the truth we are seeking is "already implicit in the mind which seeks it, and needs only to be elicited and brought to clear reflection."[4]

Thus, for Chisholm, there is nothing intellectually unrespectable about devoting a large portion of an introductory course in philosophy—or, indeed, as Chisholm himself has done, a large portion of a philosophical career—to seeking an answer to a question like 'What, if anything, justifies your belief that there is a lectern at the front of this classroom?' provided one understands this question is "Socratic and therefore not at all of the type that one ordinarily asks."[5]

---

[3] *Theory of Knowledge,* 2nd ed., p 16 The third edition of *Theory of Knowledge* (1989) is mute on the nature of the process of elicitation. See fn. 7 below

[4] *Theory of Knowledge,* 2nd ed., p. 17 Chisholm is quoting, at the end of the passage, from C. I. Lewis, *Mind and the World Order* (New York. Scribner, 1929), p. 19.

[5] *Theory of Knowledge,* p. 25; 2nd ed., p 18. Some may object that Chisholm's question is not really that similar to the question that provoked my acquaintance's complaint: (i) 'What if anything justifies your belief that there is a lectern at the front of the classroom?' The foregoing paragraphs indicate Chisholm would instead be inclined to ask (ii) 'What justification do you have for counting it as evident—for

II

I do not recall whether, or in what detail, I presented my acquaintance with this answer to her complaint. Nor do I recall whether the answer satisfied her. All I know is that the answer no longer satisfies me. It is not Chisholm's assessment of the import of the question as ordinarily understood that rankles. In ordinary contexts, requests for the justification for a belief are indeed challenges to the propriety of the belief. To accept such a challenge and fail to meet it requires the admission that the belief in question is unfounded and ought to be abandoned. As a consequence, we tend to be rather careful in deciding which challenges we are going to treat as legitimate—which challenges we are going to feel bound to accept. The decisions we make define our methodology. They determine the claims for which we will feel bound to produce justification, the claims for which we will not, and they determine what sort of justification we will feel bound to produce.

This being so, it is obvious why Chisholm should not want to be understood as urging that we enter philosophy classrooms and ask, as if in an ordinary context, "What justifies your belief that there is a lectern at the front of the classroom?" He does not mean for us to cast aspersion on the propriety of our students' belief. He does not mean for us to challenge the soundness of that methodology which sanctions their belief. It is not that our students provide poor targets at which to aim methodological advice. They come to us with many

---

thinking that you know—that there is a lectern at the front of the classroom?' And questions (i) and (ii) may not seem to be asking the same thing.

But, as William Alston has persuasively argued [in "Some Remarks on Chisholm's Epistemology," *Nous*, XIV, 4 (November 1980): 565–86, esp. 580–1], much of what Chisholm has to say about what justifies one's counting propositions as evident—as opposed to what justifies one's believing propositions to be true—is at best poorly motivated, at worst a product of confusion, and, in any event, an unnecessary embarrassment to the theory of justification Chisholm actually offers. This theory, Alston concludes, is much more successfully viewed as a response to something like the following question: (iii) 'What justification do you have for thinking that there is a lectern at the front of the classroom?'

If Alston is right (as I think he is) and Chisholm is thus most charitably read as attempting to provide us a way to answer question (iii), then there is no obstacle to reading him as I have done: as attempting to provide us a way to answer to question (i). It simply needs to be added that Chisholm is to be distinguished from other epistemologists who offer answers to question (i). he insists that only such justification that a person *has*—only such justification that a person "can know just by reflecting upon his own state of mind" (*Theory of Knowledge*, 3rd ed , p 7)—is capable of justifying that person's beliefs.

unreasoned prejudices, full of beliefs which are open to legitimate challenge and which deserve to be withheld until such time as the challenge is met. But surely their belief in the existence and location of the lectern is not among them. And just as surely the challenge that would, if accepted, have them abandon that belief until or unless they produce an appropriate argument on its behalf is a challenge that they can safely decline without fear of sinning against the strictures of good methodology.

What bothers me is the way Chisholm would reconstrue the question. To see why I am bothered, imagine that you are sitting in a doctor's crowded waiting room. The novel you have brought to read while you wait has just become interesting when the man sitting three seats to your left rises, points to you, and declares, in a voice loud enough for all to hear, "This person is a traitor to our country!" Needless to say, you are taken aback. "What, if anything, justifies your belief that I am a traitor?" you ask the man. Your accuser is calm. He explains that, although he recognizes that in accusing you of treason he has made the sort of claim that may well require justification, he cannot at present produce the justification for you. Nonetheless, he continues, he is confident that he is indeed justified in his belief that you are a traitor and, thus, he is confident that, with enough time and effort, an appropriate justification for this belief can be elicited from him.

I suspect that I know what you will want to say to your accuser. You will want to say that it is irresponsible to make a claim—and, more to the point, intellectually dishonest to believe a claim—for which one admits he is not now able to produce appropriate justification. You will want to say that it is the justification one produces that should convince him of the justifiability of the claim, not his conviction in the justifiability of his claim that should convince him that there must be a justification that can be produced for it.

You will want to say this but you will not. For you have just noticed for the first time that your accuser has in his hands the second edition of Chisholm's *Theory of Knowledge*. Indeed, you notice for the first time that everyone in the waiting room has been reading the second edition of *Theory of Knowledge* and that they all have their books open to pages 16 and 17—the very pages from which I quoted above. And you realize that they have all anticipated what you will want to say and are ready to cite Chisholm in defense of your accuser. For Chisholm has just taught them that there is nothing intellectually dishonest in believing a claim to be true, and believing oneself justified in believing that claim to be true, while admitting, at the same time, both that it is the kind of claim that may well require justification and that one cannot at present produce appropriate justification for believing the claim to be true. On the contrary, Chisholm has just taught them that this is exactly the attitude they should want to adopt toward claims when they conduct a philosophical investigation into the theory of evidence.

Something has gone very wrong here. And what seems to be responsible is a methodological precept that your fellow patients have extracted from the passages in *Theory of Knowledge* I quoted above —a methodological precept they have taken very much to heart. That precept holds that, even if you are very confident that a belief you harbor may stand in need of justification, your inability to produce a satisfactory justification need not undermine your entitlement to persist in that belief. So long as you are confident in the propriety of the belief, you are entitled to believe that, only given enough time and energy, you could elicit an appropriate justification for the belief. That is, you are entitled to believe that your inability to produce such a justification, either now or in the foreseeable future, is of no methodological concern.

It is hardly a precept that we want to be teaching our students. We do not want to be telling them that the critical evaluation of one's cherished opinions is a purely academic exercise. We do not want to be telling them that methodological hygiene requires only that one be confident that one's opinions are justified—that one need not be able to produce the justification, that one need not be able to produce any reason for thinking one's opinions are justified.

But if we are not willing to endorse this methodological precept, then we cannot endorse the way in which Chisholm would reconstrue the question 'What justifies your belief that there is a lectern at the front of the classroom?'—we cannot ask our students to do what Chisholm would have us ask of them. For Chisholm's reconstrual of the question would have us asking them to begin with a belief that they take to be justified and attempt to elicit from themselves their justification for that belief. But then we would be telling them that they are entitled to persist in their belief and to regard their belief as justified even while, engaged in the process of elicitation, they have yet to produce the justification for that belief. And this is precisely what we, fresh from the nightmare scenario, have concluded that we do not want to be telling them. It is because we do not want to have to tell them any such thing that we are unwilling to endorse the methodological precept.

III

To many, I expect that it will seem obvious that Chisholm is being unjustly accused—that his indictment on the charge of promulgating a dangerous methodological precept is the product of some unfortunate confusion. Surely there is a sense of 'justified' in which a person may be justified in believing a proposition even though she cannot produce her justification for believing it. Surely it is just this sense of justification which philosophers have had in mind when they have held that justification is required for knowledge. Surely, were it otherwise, the imposition of this requirement would require drastically revising downward our estimate of the scope of our knowledge. William Alston[6] writes that what plausibility there is to the requirement that a person be able to produce the justification for her justified beliefs is due, at least in part, to "the confusion between 'justification' in the sense of *being* justified and justification in the sense of 'showing that one is justified'. If one fails to keep that distinction in mind," he continues, "one is liable to suppose that in order to *be* justified in believing that *p* one must be *able* at least to 'justify' one's belief that *p* in the sense of showing that one is justified, i.e., exhibiting what it is that justifies one" (*ibid.*, p. 145).

But this is just more dangerous talk. If it is all right for Alston to warn you not to confuse the alleged two senses of 'justified'—if it is all right for him to explain away the plausibility of supposing that a person cannot be justified in believing *p* yet unable to produce her justification for believing *p*—then it is all right for your accuser in the nightmare scenario to do likewise. And it is hard to see how any consideration about how best to analyze '*S* knows that *p*' could possibly bear the weight of our having to say that it is all right for your accuser to dismiss your demand that he be able to produce his justification for believing you a traitor on pain of withholding that belief. The charge of confusion simply provides an apology for the unacceptable methodological precept your fellow patients extracted from Chisholm.

Of course, the mere fact that Alston fails to uncover a confusion that will save Chisholm and his project from indictment, on the charge of promulgating a dangerous methodological precept, does not mean that it is a mistake to think that this indictment is the product of a confusion. Alston's failure shows only that the task of saving Chisholm has to be gone about in some other way.

---

[6] "Level-Confusions in Epistemology," in P. French, T. Uehling, Jr., and H. Wettstein, eds., *Midwest Studies in Philosophy V: Studies in Epistemology* (Minneapolis: Minnesota UP, 1980), pp. 144–5.

What is needed, in order to save Chisholm's project from the indictment, is not a way to defend the precept your fellow patients have extracted from Chisholm's writings, but rather a way to deny Chisholm's commitment to it. What is needed is some way of saying why it is legitimate for our students to persist in their belief that there is a lectern at the front of the classroom even as they acknowledge that they cannot produce their justification for this belief, while it is not legitimate for your accuser in the nightmare scenario to do likewise with respect to his belief that you are a traitor.

It is not too hard to see what we need to say. The precept that one may legitimately persist in a belief and regard it as justified even as one acknowledges that one cannot produce the justification for that belief, we must admit, is indeed a dangerous one as ordinarily understood. It would indeed treat the task of furnishing support for beliefs as an occupation that has no more than academic importance. But, we must contend, this is not how the precept your fellow patients have extracted from Chisholm is meant to be understood. Just as the question 'What, if anything, justifies your belief that there is a lectern at the front of the classroom?' is not meant to challenge the propriety of that belief, so the precept that you may persist in a belief without (yet) being able to produce the justification for it is not meant to entitle you to disregard legitimate challenges to the propriety of your beliefs.

The impressions to the contrary are, we must claim, the product of a confusion. In ordinary contexts, questions about whether one is justified in one's belief are questions about one's intellectual responsibility in holding the belief, about one's entitlement to persist in the belief, about the propriety of one's methodology. But Chisholm is no methodologist. He is not trying to prescribe canons of rational inquiry. He means to fix on beliefs (like the belief that there is a lectern at the front of the classroom) that we already consider to have met all legitimate challenges to their propriety and only *then* to ask what, if anything, justifies those beliefs.

But then, of course, Chisholm can hardly be guilty of promulgating a dangerous methodological precept, he can hardly be guilty of sanctioning your accuser's behavior. Even by his own lights, your accuser's belief that you are a traitor has not met all legitimate challenges to its propriety. Your accuser persists in his belief, indifferent to the fact that his accusation fails to meet what he himself acknowledges to be legitimate challenges to its propriety. And there is nothing in Chisholm's project as just described to give your accuser any comfort in such a posture.[7]

IV

This having been said, we can indeed pose to our students the question Chisholm would have us pose without any fear of instilling in them any dangerous methodological precept, without any fear of disturbing our sleep with the nightmare scenario. There is, instead, a new worry. How are our students supposed to decide upon an answer?

The only source of intuitions concerning justification we have is reflection upon those canons which govern our day-to-day conduct of inquiry and criticism. It is to our practice of inquiry and criticism, in contexts both ordinary and scientific, that we turn in order to arrive at and test our conceptions of good argument and evidential support. But notice that, for the purpose of answering the question Chisholm would have us pose to our students, this source is now completely lost to us. For the canons that govern our day-to-day conduct of inquiry and criticism concern justification as it bears upon proper methodology and intellectual responsibility. And this is just the sense of 'justification' with which, we have had to say, Chisholm is not concerned. As we have had to describe his project, Chisholm would have our students begin their attempt to elicit the justification for a belief only after the belief has already met all the challenges that our canons of inquiry and criticism issue—only after all the guidance offered us by our canons of inquiry and criticism has been exhausted.

---

[7] It should be noted that, while this strategy for extricating Chisholm from the present difficulty works for all of Chisholm's epistemological writings through the 2nd ed. of *Theory of Knowledge*, it does not fare so well for the 3rd ed. There we see a radical departure. The passage from Lewis (see fn. 4), which Chisholm quotes again and again in almost twenty-five years of epistemological writing, is nowhere to be seen. The discussion of how a theory of justification is elicited—the discussion in which Lewis's quotations play a key role for those almost twenty-five years—is nowhere to be found. And suddenly (p. 5) Chisholm is writing that his inquiry into justification is designed, not to acquire self-knowledge, but rather to improve the inquirer's system of beliefs, weed out the false ones, and replace them with true ones.

Chisholm's move to the new position, apparently designed to help him fight off a rising army of externalists (see, e.g., p. 76), is an unfortunate one. After all, what market can there be for the sort of advice Chisholm would now claim to be dispensing? No one needs to be told how to arrive at a justified belief that there is a lectern at the front of the classroom. To maintain that the import of his investigation is methodological is tantamount to maintaining that it has no import at all. And this means that Chisholm's current doctrine would not have offered even an initially plausible answer to my acquaintance's complaint. She wanted to know why it was not a sham to worry for part of a semester over what, if anything, justified her belief that there was a lectern at the front of the classroom. The only answer Chisholm now makes now available, 'Because when we are done you will have important advice about how to determine the presence and location of lecterns', would rightly have made her laugh.

Thus, whatever intuitions we might have hoped to call upon through appeal to reflections upon the conduct of inquiry and the practice of criticism (such as intuitions about the viciousness of certain sorts of regresses or about the defects of circular arguments) will be the wrong intuitions. They will be intuitions about justification in the wrong sense. But what else would we have our students focus upon—what else *is* there to focus upon—to get clear on what justification, in Chisholm's (literally) ***extraordinary*** sense, is like? How, in the light of the redescription of Chisholm's project we have had to issue, can we expect our students to have the slightest idea about how to answer the question Chisholm would have us ask them?

V

The answer may seem straightforward: by simply telling them more about what sort of answer we are looking for. That, in effect, is just what Chisholm does. To put matters in our terms, Chisholm tells his readers that he is looking for an answer on which the beliefs that are justified (in the extraordinary sense of justification with which he is concerned) are just those beliefs which meet all the methodological challenges we are willing to accept. He tells his readers that he is looking for a regress of justification—unearthing ever earlier links in a (perhaps branching) chain of claims each adduced in support of its successor and terminating in the claim that there is a lectern at the front of the classroom. He suggests to them that the regress can only properly be regarded as a regress of justification if every link in the chain is either immediately justified or has an ancestor link that is immediately justified.[8]

So, if we will but tell our students what Chisholm tells his readers, our students need puzzle over our question no more. Our charge to them will be clear. They are to flesh out a view on which their belief that there is a lectern at the front of the classroom is justified—indeed, on which each belief that meets all the methodological challenges we are willing to accept is justified—by virtue of that belief's being the conclusion of an argument whose ultimate premises are immediately justified for them.

But should we want to tell them what Chisholm tells his readers? His instructions to the uninitiated result, after all, from a number of deliberate choices—none of them incontrovertible nor, indeed, uncontroverted. That we can relieve our students' puzzlement by telling them what Chisholm tells his readers may well provide our students with some comfort. But unless we can see something in the extraordinary sense of justification that underwrites the set of instructions Chisholm chooses to issue, we will be able to afford our students this comfort only at the expense of our own.

---

[8] See, e.g., Chisholm, *Theory of Knowledge*, 2nd ed., chs. 2, 4.

Consider his stipulation that what is justified in the extraordinary sense coincide with what meets all acceptable methodological challenges. It is a stipulation that certainly gives rather short shrift to skeptical concerns. After all, it dismisses from the outset the possibility that there simply is no regress of justification worthy of the name which will support the students' belief that there is a lectern at the front of the classroom. On his stipulation, any account of justification too demanding to admit the existence of a regress supporting the students' belief that there is a lectern at the front of the classroom is *ipso facto* a defective account. For any such account of justification runs afoul of the stipulation's requirement that every claim that meets all the methodological challenges we are willing to accept is to count as justified in the extraordinary sense.

Chisholm anticipates the charge that, by favoring his stipulation over the skeptic's, he begs the question against the skeptic in an arbitrary way. And, disarmingly enough, he pleads guilty. He offers but two lines of defense. The first is that the skeptic is no less guilty of arbitrariness if the skeptic champions some draconian stipulation over Chisholm's. The second is that his (that is to say, Chisholm's) stipulation has at least one virtue that its competitor's lacks—namely, that it accords with common sense. After all, we do think our students are justified in the belief that there is a lectern at the front of the classroom.[9]

To us who are searching for some intuitive grounds for giving our students Chisholm's—as opposed to some other—instructions, the *tu quoque* that constitutes Chisholm's first line of defense can offer little comfort. But Chisholm's second line of defense is likely to make many of us feel more secure. Many, and perhaps most, of us will be inclined to view the choice between defending common sense and attacking common sense to be a fairly straightforward one. We will see the accommodation of common sense to be very much the virtue Chisholm takes it to be.

---

[9] *The Problem of the Criterion* (Milwaukee: Marquette, 1973), pp. 37–8; repr. in *The Foundations of Knowing*, pp 61–75, see p. 75. Also see *Theory of Knowledge*, 2nd ed., p. 121.

But how can it possibly be a virtue that Chisholm's stipulation possesses? If common sense had anything to say about the extent of justified belief (in the extraordinary sense), there would be no need for Chisholm to issue his stipulation. Recall that the reason Chisholm needs to make his stipulation is precisely because justification in the extraordinary sense is so removed from the realm of common sense that we cannot legitimately expect our students to have any idea of what (in that sense) justifies beliefs. So Chisholm's stipulation cannot have the virtue of according with common sense. His stipulation concerns the extent of justified belief in the extraordinary sense. And that is a matter about which common sense tells us nothing.

Indeed, where common sense does reign—over our everyday canons of inquiry and criticism—there is no issue between Chisholm and his skeptical competitor. All common sense dictates is that there is no methodological challenge worthy of acceptance which impeaches the propriety of the students' belief that there is a lectern at the front of the classroom. All common sense dictates is that there is no regress of justification on pain of producing which the students should withhold belief. But this in no way touches what is at issue between Chisholm and his imagined skeptical rival. Theirs is not a dispute over the scope of justification as it bears on proper methodology. Theirs is a dispute about the scope of justification in the extraordinary sense. And, as we have noted, questions about the scope of justification in this sense have no bearing whatsoever on methodological issues. Thus, the skeptical party in the dispute is as easily able as Chisholm to accommodate the dictate of common sense. He can just as easily as Chisholm grant the authority of common sense on matters within its purview. For on the matter of dispute between him and Chisholm—i.e., on the matter of how many and what sorts of beliefs count as justified in the extraordinary sense—common sense is mute.

Chisholm then has nothing but a *tu quoque* to offer in defense of his stipulation that the class of beliefs that are justified in the extraordinary sense coincide with the class of beliefs that survive all the methodological challenges we are willing to accept: his only defense is that his stipulation is no more arbitrary than any other one might choose.

And this is but the first concession to arbitrariness. The dispute between Chisholm and the skeptic in no way exhausts the extent to which the propriety of Chisholm's instructions to the uninitiated can be, and has been, controverted. After all, Chisholm and his skeptical rival display a large area of agreement. They would both have our students look for the justification of a belief to take the form of a regress of justification, where every claim in the regress either is, or has an ancestor claim that is, immediately justified.

But there are those who would deny that a regress of justification need trace the pedigree of a belief back to beliefs that are immediately justified. Indeed, there are those who would deny that what justifies a belief need take the form of a regress at all.[10] What has Chisholm to say to these critics? What can he say in favor of seeking an account of justification in the extraordinary sense which satisfies his chosen constraints rather than theirs?

One might think that Chisholm's best move is to offer counterexamples to his rivals' proposals, to show how his rivals' proposals betray common-sense intuitions where his own proposal does not. But if such a move really had any force, we would not have found Chisholm in so desperate a position in his dispute with the skeptic. What drives Chisholm to that position is the fact that, once his project is redescribed so as to clear it of the charge of promulgating a dangerous methodological precept, Chisholm can no longer legitimately claim to be taking any position to which common sense commits us. And if Chisholm is not entitled to claim that common sense is on his side as he battles the skeptic over the extent of justified belief in the extraordinary sense, he certainly is not entitled to do so as he fights his nonskeptical rivals over the defining characteristics of justified belief in the extraordinary sense.

Thus, with common sense silenced and the intuitions provided by reflection upon our practice of inquiry and criticism exhausted, Chisholm has no more leverage against these critics than he has against the skeptic. All Chisholm can do in defense of seeking his particular sort of account of justification in the extraordinary sense is point to his rivals and utter a *tu quoque:* all he can say is that the decision to seek his sort of account is no more arbitrary than a decision to seek one of theirs.

But although this sort of defense may succeed in keeping Chisholm's rivals—skeptical and nonskeptical—at bay, it does nothing to alleviate our discomfort. On the contrary, it serves only to make vivid just how high a price needs to be paid to clear Chisholm of the charge that, in reconstruing the question 'What justifies your belief that there is a lectern at the front of the classroom?', he is promulgating a dangerous methodological precept.

---

[10] Adherents to a coherence view of justification would belong in the first group, adherents to a reliabilist view of justification to the second.

For what the concessions to arbitrariness make clear is that the extraordinary, methodologically inert sense of 'justification' with which we have had to say Chisholm is concerned, is a sense of 'justification' with which we can claim no familiarity. Common sense tells us nothing of its nature. Reflection upon the canons of inquiry reveals nothing of its contours. Even the question of whether there *are* any beliefs that are justified in this extraordinary sense is one about whose answer our intuitions offer us no guidance. The concept of justification, in this extraordinary sense, is a creation *ex nihilo*.

But then the theory of justification in this sense, far from being an attempt to answer in a systematic way to our pretheoretic intuitions, is an enterprise that is answerable to none. It is nothing more than an exercise in pure invention and pure stipulation—a game in which every move is legal and each participant plays according to her own fancy. Chisholm's project, redescribed so as to clear it of the charge of promulgating a dangerous methodological precept, becomes instead a project dedicated to uncovering the nature of a concept that has, in fact, no nature to uncover. In having so redescribed it, we have simply traded one nightmare for another.

VI

What has Chisholm done to deserve to be so haunted by the dark realm of restless sleep? What is it about his reformulation of the question that should be held responsible?

One tempting answer is that it is the grip that the foundationalist conception of justification exercises upon Chisholm which deserves the blame. After all, it is because he is in the grip of this picture that he looks for the answer to the question 'What justifies your belief that there is a lectern at the front of the classroom?' to take the form of a regress of justification. And it is because he looks for a regress of justification, because he sensibly does not want to cast aspersion on the propriety of the students' belief that there is a lectern at the front of the classroom, and because he rightly cannot expect them without much labor to be able to produce an appropriate regress that will support that belief, that he must say that they are entitled to persist in their belief and regard their belief as justified even as they seek to elicit from themselves the regress of justification that supports the belief.

And this is just what gets him into trouble. For it is precisely this entitlement—to persist in a belief and regard it as justified even as one admits that one cannot produce the regress of justification that supports the belief—that your accuser in the nightmare scenario claims on his behalf when you demand that he justify his charge that you are a traitor. It is precisely in the granting of this entitlement that Chisholm opens himself to the charge that he is promulgating the dangerous methodological tenet.

This diagnosis of the underlying cause of Chisholm's misfortune is not only plausible. It is also pleasing. The foundationalist picture that so grips Chisholm has, after all, lost its hold on most epistemologists. Other pictures, coherentist and/or naturalist in hue, command much more allegiance from those who write on epistemology today. If it is Chisholm's adherence to foundationalism that is dragging him into black depths of the dreamworld, then perhaps that just shows that there is yet one more reason to do what most epistemologists have already done. To sleep well we need only steer clear of foundationalism and head for one of the epistemologies more in vogue. But which one?

Clearly, it is not to the advocates of coherentism that we should look for relief. Their way of reading the question that caused my acquaintance such distress is in no material way different from Chisholm's. Like Chisholm, the coherentists view themselves as attempting to elicit a regress of justification for our ordinary beliefs about the world. They differ from Chisholm only in their account of the structure a regress of justification is required to exhibit. A coherentist epistemological tract will stand in very well for Chisholm's *Theory of Knowledge* in the nightmare scenario—very well indeed.[11]

In any event, coherentism is not the epistemological doctrine to which most epistemologists appear to have turned. The epistemology of the moment is clearly reliabilism. And the exponents of reliabilism do indeed view their epistemological project in a manner rather different from that in which Chisholm views his. Reliabilists hold that what justifies a belief is its being the result of a process of belief formation that reliably produces true beliefs. It matters not at all whether the person who provides a harbor for the belief believes, or has any reason to believe, that the belief she harbors enjoys this fortunate provenance. Thus, for reliabilists, the search for what justifies the students' belief that there is a lectern at the front of the classroom does not lead, as it does for Chisholm and the coherentists, to a theory of evidence. What justifies their belief is no implicit chain of reasoning—either linear or looped. It is, rather, a fact about the process through which that belief was formed—a fact of which the students might well be unaware.

---

[11] This is nowhere more clear than in Laurence Bonjour's *The Structure of Empirical Knowledge* (Cambridge: Harvard, 1985), pp. 151–2, where he writes: "Thus the basic claim of my coherentist position (and indeed of the analogous claim for most or all nonexternalist versions of foundationalism) must be that careful reflection on actual cases will reveal that these elements [the inferences and belief systems necessary for justified belief] are tacitly or implicitly involved in the actual cognitive state of a person who has empirical knowledge, even though he does not bring them explicitly to mind and indeed would normally be unable to do so even if explicitly challenged" (p. 152).

Consequently, the nightmare scenario that exploits Chisholm's response to the sort of complaint issued by my acquaintance—a nightmare scenario that might equally well have featured the response of a coherentist epistemologist—need raise no goosebumps on a reliabilist's skin. Part of what makes the nightmare scenario nightmarish is that your accuser admits that his accusation is the sort of claim that requires justification, yet he—along with everyone else in the waiting room—is entirely unperturbed by his inability to produce the appropriate justification. The reliabilist, however, refuses to acquiesce in your accuser's admission. For the reliabilist, there is no claim for which a person must have grounds if she is to be justified in believing it. All that is required is that her belief be the result of a reliable true-belief-forming process. (It may happen that the process produces grounds for her belief, but it need not.[12])

VII

But it would not be wise to count just yet on a restful night's sleep. For imagine that it is the reliabilist's epistemological doctrine that is fresh on the minds of your fellow inmates in the doctor's waiting room as you demand from your accuser the justification he has for his outrageous claim. It is true that you need no longer expect that he will admit that his is the sort of claim that requires justification, admit his inability to produce this justification, yet insist that he is entitled to persist in his belief in your treachery and regard that belief as justified just the same.

Rather, you can expect that he will find your demand rather quaint. You can expect him to explain to you that the notion that one needs to be able to provide grounds for one's beliefs in order to have justified beliefs has long been discredited. What makes a belief justified—what makes his belief in your treachery justified—is the fact that it is a product of a reliable true-belief-forming process. His inability to provide grounds either for his claim that you are a traitor or for his claim that his belief in your treachery was produced in the way he described, you can expect him to continue, in no way indicates that he is not justified in believing these claims. Grounds are simply beside the point. You can also expect that, all the while, your fellow patients will be nodding in agreement.

---

[12] See, e.g., the reliabilist theory of justification defended by Alvin I Goldman in "What is Justified Belief?" in *Justification and Knowledge*, George S. Pappas, ed., (Dordrecht: Reidel, 1979), pp. 1–23

The nightmare is all too familiar. Once again, your fellow patients have extracted from a reputable epistemological doctrine a methodological precept that would sanction your accuser's reply to your demand that he produce grounds for his outlandish accusation. Once again, it is a dangerous methodological precept. And, once again, it looks as if we cannot ask our students to endorse the reputable epistemological doctrine from which the dangerous methodological precept has been extracted without thereby promulgating that precept.

But this time it is reliabilism, not foundationalism, that is to blame. It is from reliabilism that your fellow patients have extracted the precept that, since it is the reliability of the process producing a belief which renders the belief justified, and since one's being able to meet legitimate methodological challenges to a belief is neither necessary nor sufficient for its being produced by a reliable process, one is free to dismiss any such challenges; to the question of whether one's belief is justified, one's ability to meet such challenges is irrelevant. This methodological precept is no less dangerous than the one extracted from Chisholm. And it looks to be no less implicit in the reliabilist account of justification than the previous precept was in Chisholm's.

Not that there are no responses available to the reliabilist. Like Chisholm, he can claim that his doctrine is not meant to have methodological import,[15] and thus that his answer to the question of what justifies a belief can provide no comfort to your accuser. He can even claim, if he wishes, that the question he means to answer arises only after all legitimate challenges to the belief have been satisfactorily met (or, that the question arises only for beliefs against which no legitimate methodological challenges can be raised). But these responses can get him no farther than they got Chisholm. Once the reliabilist denies that his use of 'justified' is to be understood in the ordinary way, he relinquishes his right to call upon ordinary intuitions concerning the nature and extent of justified belief in support of his theory of justification. And without recourse to such intuitions, his theory can be nothing more than an exercise in pure stipulation.

---

[15] As Goldman does in *Epistemology and Cognition* (Cambridge: Harvard, 1986), pp. 25–6, and in "The Internalist Conception of Justification," in *Midwest Studies in Philosophy V: Studies in Epistemology*, p. 28.

VIII

There are some writers, reliabilist and nonreliabilist, who will undoubtedly protest this description of their predicament. They will claim, as we have already seen Alston do, that the distinction between uses of 'justification' to which they appeal and which I have so deprecated is, in fact, underwritten by our ordinary intuitions. But this claim, of itself, cannot reverse their fortunes.

What anchors our intuitions about the nature of justification as we ordinarily understand it is the role ascriptions of justification play in the conduct of inquiry and criticism. Questions about justification, as the term is ordinarily understood, are questions about the propriety of methodological commitments and the propriety of adopting (retaining) the beliefs these commitments sanction. These are consequential questions. To conclude that a belief has not the propriety we supposed is to demand that it be denied the credence it has hitherto been accorded. It is the fact that we have an actual practice of inquiry and criticism at stake when such questions are raised which places constraints on what answers we can, in all honesty, take seriously.

To uphold the claim that ordinary intuitions support the existence of some other sense of justification, something comparable needs to be said about what anchors intuitions about justification in this other sense. Something needs to be said by way of describing the endeavor in which ascriptions of justification in this other sense are supposed to play a role—something to convince us that our intuitions do indeed place substantive constraints on theories of justification in this other sense. But what is there to say?

Some have said that it is our intuitions about knowledge that anchor our intuitions about the concept of justification in which they are interested. It is justification, in the sense that justification is necessary for knowledge, into whose nature they mean to inquire.[14] But this is an appeal we can safely set aside. If Gettier has taught us anything, it is that, after we have determined that *P* is true and decided that an investigator has met all legitimate methodological challenges to her belief that *P* (i.e., all challenges she must meet on pain of having to withhold belief that *P*), the question of whether her belief counts as knowledge is still open. One lesson, then, is that, whatever is at stake in how we answer that open question, it has nothing to do with the propriety of the manner in which the investigator has conducted her inquiry; its propriety has already been settled. But that is just to say that we are free to answer the question of whether her belief counts as knowledge however we like—we are free to adopt any position concerning the nature and extent of propositional knowledge—without fear that we will have to pay a price for our stand, that there will be something in our conduct of inquiry or anything else that will have to be trimmed to fit. And that means that calling upon intuitions about knowledge to anchor intuitions about justification cannot deliver the help needed. For it means that intuitions about knowledge are themselves adrift from any and all human endeavors that might conceivably constrain them.[15]

---

[14] For such an appeal, see Alston, "Level Confusions in Epistemology," p. 144; Goldman, "The Internalist Conception of Justification," p. 29.

The only other gambit of any promise seeks to exploit an apparent analogy between the epistemic and moral realms. When contemplating an agent who must make a moral decision, we can ask 'How ought that agent to act?' There are at least two different things we can mean. We can mean to be asking: 'What does the correct moral theory require the agent to do in her circumstances?' Or we can mean to be asking: 'What does the correct moral theory require the agent to do given what the agent takes to be her circumstances?'[16] Meant in the first way, the question asks which among the available courses of action is *objectively right*—if act utilitarianism is correct, which of the available courses of action will maximize happiness. Meant in the second way, the question asks which among the available courses of action is *subjectively right*—if act utilitarianism is correct, which of the available courses of action would maximize happiness were the world the way the agent takes it to be.

Objective and subjective rightness often fail to be satisfied by the same acts. An act-utilitarian agent who has misleading information about her circumstances may well perform an act thinking it will maximize happiness when, in fact, it will not. If act utilitarianism is the correct moral theory, her act will then have been subjectively right but objectively wrong. It is clear from this case that the latter assessment is far removed from regulative concerns. The verdict that her act was objectively wrong attaches no taint to the agent—it is compatible with the verdict that no one in her circumstances could have done any better. It is in our assessment of the subjective rightness of the agent's act that we express regulative concern—our concern about whether she did the best she could in her circumstances, whether she ought to take steps to improve her performance. But the two assessments, subjective and objective, are this much connected: an agent's act is objectively right if and only it is an act she would be prepared to perform were she to know and act upon all the moral and nonmoral facts relevant to her decision.

---

[15] I offer a fuller discussion of these matters in "It's Not What You Know that Counts," this JOURNAL, LXXXII, 7 (July 1985): 350–63.

[16] A third thing we can mean, but which we can ignore for the present discussion, is: "What is the agent required to do given what she takes to be the correct moral theory and what she takes to be her circumstances?" For discussion of this way of taking the question, see Hilary Kornblith, "Naturalizing Rationality," in Newton Garver and Peter H. Hare, eds., *Naturalism and Rationality* (Buffalo: Prometheus, 1986), pp. 115–33.

Our intuitions, it has been claimed, support an analogous distinction in epistemology. Faced with an agent harboring (or about to harbor, or about to abandon) the belief that *P*, we can ask 'Is the agent's belief justified?' Again, it is claimed, we can mean at least two things. We can mean to be asking whether the agent's belief is *objectively* justified or whether it is *subjectively* justified. And again, the story goes, it is only meant the second way that the question expresses a regulative concern—a concern about whether the agent has met all the challenges to her belief that the proper methodology demands. When meant the first way, the question is nonregulative, devoid of methodological import. And, it is claimed, it is to our intuitions about justification in this first, objective, sense that the reliabilist theory of justification means to answer.[17]

But the claim cannot be sustained. Suppose we let the analogy between the moral and epistemic realms guide our intuitions. We should then expect, not only that the two ways of assessing justification should express different concerns, but also that they should exhibit the same sort of connection we see in the moral case. We should expect that an agent's belief will be objectively justified if and only if it is a belief she would hold were she to know and act upon all the epistemic and nonepistemic facts relevant to the adoption (retention) of the belief. But, given that truth is a central aim of inquiry, one key fact relevant to the adoption (retention) of an agent's belief must be whether the belief is true. So, following the analogy, we should expect that an agent's belief will be objectively justified if and only if it is a belief she would hold if she knew the truth value of the propositional object of the belief. And from this it follows that all and only true beliefs are objectively justified. But this is a condition that no reliabilist (or foundationalist or coherentist) appears even slightly inclined to take seriously.[18]

---

[17] This appears to be like Goldman's view. See his *Epistemology and Cognition*, p 25 (see fn. 4); and "The Internalist Conception of Justification," pp 28–9.

One could, of course, contrive to loosen the relation between the criteria of subjective and objective moral rightness, or to loosen the analogy between the moral and epistemic realms, or both. But, it is precisely the tightness of the relation of objective to subjective moral rightness—the fact that the assessments of objective moral rightness are just assessments of subjective moral rightness made in the light of knowledge that may be unavailable to the agent—that lends the concept of objective moral rightness *its* intelligibility, *its* anchor in matters of practical import. And it is the tightness of the analogy between the moral and epistemic distinctions which is supposed to lend the concept of objective justification its intuitiveness. To engineer this sort of escape from the consequences of the analogy drawn is to forego any intuitive cachet the analogy might have afforded the enterprise of constructing a theory of justification, where 'justification' is understood in any sense out of the ordinary.

Thus, if objectively justified belief is not just true belief, then the appeals to intuitions about knowledge and about objective justification offer no relief. Unless something has been overlooked here, there is simply no reason to think that our intuitions give us purchase on any extraordinary notion of justification at all.

IX

The tempting, happy diagnosis of the cause of Chisholm's misfortune is, thus, mistaken. It is not Chisholm's adherence to foundationalism that explains his encounter with the fearful world of dreams. Chisholm's coherentist and reliabilist counterparts have to contend with much the same nightmares. Rather, all the nightmares have their origins in one and the same assumption—an assumption that Chisholm shares with his coherentist, reliabilist, and even skeptical colleagues. The assumption is that, even after our students have satisfied us that they have done everything that intellectual responsibility and methodological hygiene demands of someone who believes that *P* (in their circumstances)—that they have met all the challenges that need to be met on pain of withholding the belief that *P*—we can still sensibly say to them, "Yes, yes, but you haven't yet finished saying what (if anything) justifies your belief that *P*." The assumption is that there is something more to say.

---

[18] John Pollock offers much the same argument (albeit in service of a somewhat different cause) in "Epistemic Norms," *Synthese*, LXXI, 1 (April 1987): 61–95; see pp. 77–8.

But when we have examined what more there is supposed to be left to say, it has invariably turned out to be something that, read flatly, offends against uncontroversial tenets of sound methodology and, read with the aid of the sort of distinction required to prevent the offense, succeeds in so doing only by employing some sense of justification about which we can claim no intuitions at all. We could continue our survey of what more there is supposed to be left to say, in the hope of finding some account that will succeed where foundationalism, coherentism, reliabilism, and skepticism have failed. But there would be no point.

We have already learned that, on pain of losing our intuitive bearings entirely, we have no choice but to read our talk about justification flatly. But, read flatly, we cannot claim that our students have not yet finished saying what (if anything) justifies their belief that *P* without thereby accusing them of having failed to have done all that methodological hygiene and intellectual honesty demand of someone who believes that *P*. Thus, once we have conceded that our students have met all legitimate methodological challenges to the propriety of their belief that *P*, we have conceded that they are finished saying what justifies their belief that *P*.

Small wonder then, that, in attempting to tell our students how to say something more, Chisholm and his coherentist, reliabilist, and skeptical competitors should end up saying things that, read flatly, run afoul of methodological precepts. Small wonder that there should be no hope that we might shop around and somehow find a better way to instruct our students on how to say something more. There is nothing more to say. There is nothing more to say in response to the question 'What justifies your belief that *P*?' than intellectual honesty and methodological hygiene demand that we be able to say on pain of withholding the belief that *P*.

And, if so, my acquaintance's complaint turns out to have been indeed well-taken. She found her instructor's challenge to say what, if anything, justified her belief that there was a lectern at the front of the classroom to be a sham, a display of intellectual dishonesty. Those are harsh accusations, but she turns out to have been this far right: the question admits of no interpretation under which it deserves, or can receive, a serious answer.

Those familiar with Chisholm's corpus will recognize the dialectical significance of this result. The passages I quoted at the outset were not, after all, written in response to my acquaintance's complaint. Chisholm wrote them in response to philosophers of pragmatic and/or Austinian and/or Wittgensteinian bent who had argued that asking someone for the justification of an everyday perceptual belief is perverse and/or an abuse of language.[19] Chisholm's strategy, recall, was to admit the force of their objections to the question as ordinarily understood but to suggest that there is room to understand the question in a different way. Chisholm's ambition was thus to undercut the pragmatic/Austinian/Wittgensteinian critique to which my acquaintance simply gave a naive voice.

What we have seen is that Chisholm's ambition cannot be realized, his strategy cannot succeed: there is no room to understand the question Chisholm wants to ask in anything but the ordinary way. The perversity that my acquaintance saw in the question her instructor asked by way of drawing his class into the concerns of contemporary epistemology—the perversity to which Peirce, Austin, and Wittgenstein tried to draw attention—cannot be finessed away. The question that Chisholm wants to ask—the question that his foundationalist, and his rivals' coherentist, reliabilist, and even skeptical theories of justification would answer—is fundamentally misguided.

Were it just the question that deserved the blame, this neo-pragmatic/Austinian/Wittgensteinian storm might yet be weathered. After all, the mere fact that the theories of justification we have canvassed can be regarded as attempts to answer a misguided question—even the fact that at least some of these theories have been regarded from the outset as attempts to answer this question—hardly means that they must be so regarded. Were it just this question that deserved repudiation, we might still be able to view these theories in some other, more favorable, light—as the products of some other, better motivated enterprise. And if so, foundationalism, coherentism, reliabilism, and (perhaps) even skepticism might survive the foregoing critique and retain their claim to be competing theories of genuine interest.

But if there really were something about which foundationalism, coherentism, reliabilism, and skepticism could properly be said to be competing theories, there would be no reason to cast aside Chisholm's question. We could simply interpret Chisholm's question as that question to which these theories offer competing answers. The reason we have deemed Chisholm's question misguided is that we have learned that, unless we want to say that these theories are promulgating dangerous methodological precepts, we cannot but conclude that there simply is no pretheoretic concept of justification about whose nature they can properly be said to differ.

---

[19] See the passages cited in fn. 1 and 2

It is not Chisholm's question, but rather the vision of the theory of justification which he shares with his coherentist, reliabilist, and skeptical confreres which is the problem. Theirs is a vision of a pure theory of justification, separated from all that would make it methodologically potent, cleansed of the concerns proper to the realm of the ordinary. The problem is that a theory of justification thus purified, a theory of justification deprived of any role in methodology or the conduct of inquiry and criticism, is a theory that divorces epistemology from the very practices that furnish it with its only source of intuitive constraint. It is epistemology on holiday.

We need not fear that, in recognizing this problem and rejecting the vision of epistemology that is responsible, we will have turned our back on the question of what justifies belief. If there is nothing more to say in response to the question 'What justifies your belief that *P*?' than methodological hygiene and intellectual responsibility demand we be able to say on pain of withholding the belief, then the question that deserves our attention is 'What exactly do methodological hygiene and intellectual responsibility demand of us?'

There is in philosophy a noble tradition of theorizing about the nature of justification which addresses just that question—which seeks to clarify, to criticize, to improve the conduct of inquiry and criticism. It is a tradition exemplified in the seventeenth century by Descartes (*Rules for the Direction of the Mind*), in the first half of our own century by the authors of logical empiricism, and in contemporary writing (I would argue) by advocates of the Bayesian approach to epistemology and the philosophy of science. And this is to mention but three instances in which philosophers have undertaken to confront deep methodological issues and evaluate the way in which inquiry is properly to be conducted. Much more of the very same sort of work beckons us today. Which is to say that the theory of justification can survive very well the loss of the idle question in whose name it has journeyed so far and to such little effect. Epistemology need not stay on holiday; it has a job waiting at home.

MARK KAPLAN

University of Wisconsin/Milwaukee

BOOK REVIEWS

The Right to Private Property. JEREMY WALDRON. New York: Oxford University Press, 1988. viii, 470 p. Cloth $59.00.

The Right to Private Property displays intellectual and literary virtues not often found together—boldness and sobriety, sharp focus, and broad sweep. Aware of the diffuseness of the topic of property rights, Jeremy Waldron has chosen to focus narrowly enough to ensure intellectual discipline. He concentrates only on "rights-based" arguments for private property, leaving to one side and another occasion utilitarian arguments, arguments from the preservation of distinctively "republican" political virtues, and "Aristotelian" natural-law arguments from the fostering of desirable sorts of character. By the same token, he sticks to arguments over private property, and even where these lead, as they do with Proudhon, to the conclusion that "property is theft," he does not speculate about what forms of common or public ownership a rights-based approach would license.

Yet, the first hundred and twenty pages of *The Right to Private Property* are devoted to the two largest issues in discussions of property rights, namely, the nature of property rights and the nature of rights-based arguments; and even though the discussion of rights-based theories concentrates on a handful of authors—Locke, Robert Nozick, Kant, Hegel, and Proudhon—the conclusions toward which the argument drives are of the broadest. Reversing the trick of many authors, who offer as a conclusion what is essentially an introduction, Waldron spells out his own conclusions in the first couple of pages.

Starting from the rhetorical claim that property is a human right, he asks "whether there are any good right-based arguments for private property." A right-based argument is understood in much the way Ronald Dworkin understands it, that is, as an argument from an *individual* interest to the conclusion that the interest is of such importance that government is bound to promote it. A right-based argument for private property is thus an argument to the effect that governments are obliged to maintain a system of private property as opposed to communism or some other alternative. Such arguments come in two forms. The Hegelian argues that owning property contributes immensely to the ethical development of the individual, the Lockean that those who have acquired the appropriate interest in an object, by labor or legitimate transfer, have an interest that governments are obliged to protect.

0022-362X/91/8803/155–9

The Hegelian argument thus presupposes a general right, the Lockean a special right. Waldron's argument is that the Lockean theory of acquisition, like the refurbished version offered by Nozick, is unconvincing, because there is no plausible justification of the transition from mixing one's labor to acquiring what amounts to "full liberal ownership"; nor is there any plausible justification of the obvious alternative, the argument from "first occupancy." Locke emerges with rather more credit than Nozick, because he assumes a background of a general right to subsistence which constrains the acquisition of would-be owners of property, and Waldron argues with considerable plausibility that "no theory of the kind that Nozick intimates could possibly be made acceptable in the absence of a back-ground general right to subsistence."

In the case of Hegel, the issues are different. Hegel demands that all adult persons be property owners; the argument—that only thus can they show the responsibility and autonomy that define moral agency as we conceive it—is tortuous and difficult, but engaging. Waldron does not want to claim more success for the Hegelian enterprise than he can demonstrate, but agrees that it establishes some connection between personality and ownership; perhaps more to the point, the Hegelian tradition has done it by paying attention to social and political phenomena that appear in libertarian writings only at the level of rhetoric.

Once we accept even so much of a Hegelian perspective, interesting and difficult issues abound. The most obvious is the question whether the regulation and redistribution required to ensure that *everybody owns property* will not so cut into the rights of owners to do as they choose with their own that nothing resembling private property is left. Waldron's view is that most people who claim that this would be the result of taking Hegel seriously just go too fast. Certainly, it might be true but, on the face of it, it need not be. There are numerous different conceptions of private property, and by no means all of them are inconsistent with a good deal of regulation and redistribution. Nonetheless, the egalitarianism implied by taking Hegel seriously cannot be absolute. Acknowledging the merits of private ownership is inconsistent with the suppression of all inequality. More contentious is the conclusion that acknowledging the right to property is equally inconsistent with accepting the degree of inequality manifested in capitalist economies where a few people have lots of property and lots of people have none whatever.

As this summary suggests, Waldron is unafraid of contentious conclusions. What it cannot suggest is the continuous stream of interesting and novel argumentation on the way to those conclusions. Waldron's discussion of the definition of private property, for instance, is a splendid example of brisk common sense allied to considerable subtlety of analysis. It helps that he is a lawyer as well as a philosopher, and can bring a great variety of legal material to bear on our first intuitions about the essence of private ownership. Acknowledging that many writers have despaired of defining property, for the not wholly disgraceful reason that the objects of ownership, the nature of their owners, and the legal relations between owners and what is owned have varied very widely, Waldron insists that his subject cannot properly be discussed without some pretty clear account of just what private property amounts to. Following H.L.A. Hart and John Rawls, he suggests that variations in the various dimensions of ownership can be handled by thinking of different conceptions of the one concept of private property.

As to the core notion of property, he starts from the thought that property rules are rules governing access to and control of material resources. The now famous instance of why the rule forbidding me to stick my knife in your chest seems not to be a rule about property, even though it governs my use of a material object, is easily handled; my *ownership* of the knife is a matter of the distribution of access to and control over the knife, but the rule forbidding me to stick it, or any other knife, in your chest is not. The difficulty presented by the English Common Law's doctrine of estates, which held that what the owners of land really owned was an incorporeal estate *in* the land rather than the material stuff itself, is set aside; strictly speaking, there was no private ownership of land in English feudal law. Talk of incorporeal estates is just a way of talking about a form of collective ownership.

Applied generally to the phenomena of the ownership of shares, copyrights, mortgages, and reversions, Waldron's approach is to stick firmly to the idea that property rules govern access to and control over material resources, and that where, as in a modern economy, we have widespread corporate ownership of the material resources themselves and private ownership of shares in the corporation, we should say that the man who owns *shares* does not own material resources, even though his wealth does indeed lie in the rules that determine access to those material resources. This is thoroughly sensible, but leaves me at least unsure just what the upshot is. Is 'ownership' equivocal as between owning things and owning claims on things, are such shares as I may own *not* my property, or what? The problems that lead so many of us to say that ownership is whatever counts as ownership in a given legal system have not entirely been exorcised. If the definition of property causes trouble, the privateness of private property does not. A property regime is a private-property regime when it puts all the incidents of ownership into the hands of nameable individuals. The contentious conceptual issues have to do with how many of these incidents and which of them it takes to make it a system of private ownership, on whichever of those two words the stress is taken to fall.

The discussion of particular writers which follows these preliminaries is continuously interesting. Devotees of Locke might complain that Locke is too closely assimilated to Nozick, and admirers of Hegel might wince occasionally when Waldron presses on a favorite equivocation in the master's exposition of the individual's need to manifest his will in ownership, but overall the discussion is a model of fair-minded exposition and criticism. All the same, this is well-tilled soil, as Waldron generously acknowledges, both in the text and in his footnotes. Although Waldron's treatment of the subject is by some way the best now in print, he has had a bit of help from those who came before. Let me conclude, therefore, by mentioning one writer whose work has never been treated with the respect that Waldron accords it, and where Waldron turns all one's expectations upside down.

In twenty pages, Waldron rescues Proudhon from a century and a half of obloquy. Most twentieth-century readers who know anything at all of Proudhon know him only as the author of the famous observation that "property is theft," and thus as the target of the obvious putdown that theft presupposes property since nobody can steal what nobody owns. In Waldron's account, Proudhon simply held that we all have an equal general right to become an owner of property. This supports a positive view of the morality of proprietorship in some shape, but also gives sense to the claim that anyone whose holdings make it impossible for another to become a property owner has deprived him or her of their property, and is therefore guilty of theft.

Kant had earlier come close to such a claim when attacking feudal obstructions to the easy alienability of landed property, but Proudhon was more impressed by the thought that, if we take seriously the right of everyone to equal liberty, *and* the fact that that liberty demanded property for its expression, we must curtail the rights of existing owners for the sake of those otherwise left propertyless. What is appealing about Waldron's treatment of the argument is the deftness with which he distances himself from Proudhon's own formulation, which does indeed suggest that in curtailing the usual rights of owners we are *abolishing* property, and then uses Proudhon's insight as the basis for an egalitarian account of the distributive implications of a general right to property. It is not surprising that Waldron draws one of the epigraphs to *The Right to Private Property* from Proudhon. "Whenever a person is born, the others must crowd closer together"—the haves cannot simply insist on their rights of ownership against the right to liberty and the welfare needs of the have-nots.

ALAN RYAN

Princeton University

Fact and Method. RICHARD W. MILLER. Princeton: Princeton University Press, 1988. 614 p. Cloth $65.00, paper $14.50.

Philosophy of science often appears as a subdivided estate, its products the work of specialists. Publications, whether articles or books, tend to focus on one or another of the subtopics such as explanation, confirmation, or realism. In this context, it is good to see a new book such as Richard Miller's *Fact and Method,* in which all the major issues are treated with the care and coherence to show their close interdependence. This is by no means a piecemeal survey of the philosophy of science; rather, it is a forceful and cogent review of scientific methodology which weaves a tough thread of relevance through the analysis of explanation, confirmation, and realism.

Fact and Method, with its ambitious goal of replacing positivism as the dominant perspective on science, provides a clear and accurate atlas of the discipline of philosophy of science. As the author argues in detail against the positivist account and in favor of his replacement, we see clearly why the analyses of explanation and confirmation are essential groundwork for an argument for realism, the author's dominant cause. Accurate descriptions of these scientific activities function as the evidence for claims about the appropriate level of justification. Thus, with the author's demonstration that the complete argument for realism must be "book-length," he acknowledges the need to understand what science does before we can adjudicate how much of it is believable. And the effect of this good strategy is a philosophy-of-science book with a breadth of perspective to make it readable and informative both for the general philosophical audience and for philosophers of science.

The case against positivism begins with an account of the failure of the covering-law model of explanation. This is the core of the posi-

0022-362X/91/8803/159–62

tivist's approach to explanation and it fails to describe adequately that scientific activity because it does not attend to the causal aspects of the phenomena to be explained. A covering-law explanation shows why an event was to be expected but not necessarily what brought it about. Capturing the important features of scientific explanation, Miller argues, requires analysis as, "an adequate description of underlying causes" (60). This demands a clear concept of what it is to be a cause and an indication of how we are to recognize causes. Miller responds to this requirement by treating causality as a "core concept." Our understanding of a core concept (others being number and work of art) is based on a diverse pool of exemplars and procedures for extending the concept so that other examples can be judged as being relevantly like the exemplars. There is no general analysis of the concept. Specifications of causality will be discipline-specific in the sense that the sciences themselves must specify the exemplary causes and extension procedures. If theories of interactions change within science, so too will the concept of cause. Thus, the adequacy of a causal account, that is, the nature of explanation, is a scientific, not philosophical decision.

This reveals the principle methodological tack in *Fact and Method*, a counsel against general, a priori (philosophical) accounts of science and an emphasis instead on case-specific (scientific) principles. The pervasive flaw in positivism is its reverence for generality and its insistence on rules of scientific reasoning which span the sciences. This is the germ of its inadequacy. Miller's demonstration against this kind of generality is supported by a wide variety of examples wherein each scientific discipline fills in its own causal rules and thus sets its own explanatory guidelines. The examples in this case and throughout the book show a rich diversity of sciences, both social and natural. This dimension of breadth is another strength of the book.

The causal account helps us understand a necessary feature of explanation. But more epistemological work needs to be done in weeding out unjustified assessments of cause. This is the task of confirmation. Miller's diagnosis of positivist accounts of confirmation (hypothetico-deductive and Bayesian models alike) finds the same deficiency in their blindness to causal factors. The remedy is to cast confirmation of a hypothesis as, "fair causal comparison with its rivals" (155). That is, a hypothesis is confirmed by explicit comparison with actual alternatives, arguing from principles shared by all sides. One accepts the hypothesis that is entailed by the best causal account of the data and our beliefs about them. The evaluation covers not just the account of how the phenomena in question arise, but also the story of the interactions by which we observe the phenomena and how we come to entertain the belief. Miller provides mainstream examples (episodes in confirmation of Darwin's theory and Newton's) to illustrate how this model of confirmation maps accurately onto real science.

This coherent treatment of explanation and confirmation lays the important foundation for Miller's argument for realism. Since acceptance of a hypothesis is linked by the account of confirmation to the causal properties of the entities involved, it is also linked to truth. That is, the virtues that mark a hypothesis as the best explanation are truth-conducive virtues and not simply psychological (the phenomenon was to be expected) or pragmatic (the hypothesis is the simplest). A causal explanation has the existential claim built in, in the sense that, if you confirm the explanation involving entity *x*, then you confirm as well that *x* exists, since "what does not exist cannot cause" (382).[1]

This neat link between best explanation and truth seems too good to be true, and it is. If the account of explanation and confirmation is to carry into the fray of realism sufficient momentum to change skeptical minds, it will require significantly greater epistemological precision in describing causality. The fuzzy notion of cause as a core concept will suffice in the argument that explanation requires a causal account. But this half-analyzed concept does not provide the necessary link between our ability to recognize a cause, based on a dynamic, topic-specific core of exemplars and extension procedures, and truth in causality. The specific realist case that entity *x* exists relies not only on the truth of the conditional that, if *x* is cause, then *x* exists, but also on our ability to establish the truth of the antecedent. It is unanswered doubt about this ability which at this point leaves antirealism as an option.

But Miller's case for realism is a much richer and more detailed argument than I have indicated so far. He is specifically opposing an antirealism of that dodgy form which maintains only that an attitude of agnosticism toward the existence of unobservables is always a reasonable option. The rebuttal to this comes through specific cases (as it must, to be in line with the methodological opposition to generality) in which "topic-specific truisms" leave the scientist no reasonable alternative but to believe the particular explanatory theory. Examples of such principles that "can move all reasonable people" (469) are given for the theory of microbes and for atomic theory as an explanation of Brownian motion. In the latter case, it is simply unreasonable to deny that the acceleration of the particles is caused by external forces. Failure to endorse such a truism is failure to participate in physics. Believing such a truism is in this sense a criterion of rational participation in science.

---

[1] This is much like Nancy Cartwright's analysis in *How the Laws of Physics Lie* (New York: Oxford, 1983).

In this way, philosophy learns from science itself the standards for what is and what is not reasonable to believe. Scientists, that is, "have something to teach philosophers" (471) about realism. It will, of course, be a convincing lesson only for those who are predisposed to respect the scientist's judgment of believability. Miller acknowledges (384) this presumption of respect for science but does not consider it a serious weakness in the case for realism. But the two prerequisites, this respect of science and a definitive epistemology of causality, puts *Fact and Method* in place with most other voices in the cause of realism, compelling only to realists. Miller's contribution, though, is among those most worth hearing since its book-length argument, by weaving explanation and confirmation into a complete case, nicely demonstrates how the argument ought to go. But book-length or not, this argument is unlikely to convert the nonbeliever since its points of fuzziness will always turn out to be in the critical view.

And there is irony in Miller's methodological persuasion, which he shares with many others, to direct philosophy of science toward the specifics of science and away from the generalities of philosophy. As logical positivists marshalled metaphysical arguments to ban metaphysics from philosophy of science, Miller presents a general account that is intended to pre-empt generalizing. The causal account of explanation and confirmation avoid the positivist's generality, he claims, because it "describes jobs that must be done, but not the content of the principles that do them" (199). But clearly the description is meant to apply generally in the sense that all explanation is to speak of causes. Much like the covering-law model, Miller's account prescribes a general structure that leaves room for specific content. The real difference is that the positivists tend not to fill the general model with specific principles. They think that general claims complete the understanding of science. For Miller, on the other hand, we do not fully grasp what it is to be an explanation (for example) until we fill the general form with specific content. The persuasion then should not be so much against generality as *for* specifics. Generalizing, after all, is the accomplishment of understanding. Just as science understands the world through unification and generalization, philosophy understands science with similar style. So it is true: scientists do have something to teach philosophers.

PETER KOSSO

Northern Arizona University

Explaining Science: A Cognitive Approach. RONALD N. GIERE. Chicago: University of Chicago Press, 1988. xxi, 321 p. Cloth $34.95, paper $17.95.*

Most recent discussions of problems in the philosophy of science take a highly idealized view of scientists. Insofar as cognitive agents can be discerned at all behind the staples of philosophical analysis—the reconstructed theories, the confirmation functions, the covering-law explanations, and so forth—they are logically omniscient systems with infinite memory, who make instantaneous responses to stimuli and who act in isolation from one another. For many other students of the scientific enterprise, particularly historians and sociologists of science, the philosophical depiction of the scientist is so idealized as to lose any contact with the phenomena. Thus, in at least some quarters, philosophy of science is dismissed as a scholastic irrelevancy, a cluster of intricate games pursued with odd obsession by some apparently highly intelligent people.

Ronald Giere hopes to remedy this unfortunate situation by adopting a "cognitive approach" to science. He proposes to use ideas from contemporary cognitive science to reformulate some philosophical problems about science and to reconnect philosophy of science with the research of historians and sociologists. Left to their own devices, historians and sociologists pursue philosophy by other means, and, Giere believes, some of their philosophizing, particularly the flirtations with relativism, needs critical scrutiny. The cognitive approach is not only to correct the scholastic excesses of philosophy but to trim the ambitions of sociology as well.

Giere begins with two chapters that present his ideas about the current state of science studies and of epistemology generally. He offers thumbnail sketches of philosophical naturalism, of evolutionary epistemology, of logical empiricism, and of the work of Thomas Kuhn, Bas van Fraassen, and such contemporary sociologists as Barry Barnes, Harry Collins, and Bruno Latour. This is followed by chapters in which he formulates his own version of the semantic conception of scientific theories, using it to specify a position that he calls "constructive realism." The central idea is that scientists typically (and correctly) contend that the world is similar to specified models in specified ways and to specified degrees. The realist emphasis comes in with the contention that the ways in which models are

---

* I am extremely grateful to Gary Hardcastle for some valuable discussions and his constructive comments on an earlier draft.

0022-362X/91/8803/163–7

akin to reality are not confined to agreement on observables: typically, scientists will claim approximate agreement in a far wider variety of respects.

In his fifth chapter, Giere attempts to defend constructive realism by appealing to his experiences at the Indiana University Cyclotron Facility (IUCF). His targets include empiricists (primarily van Fraassen) and social constructivists (such as Karin Knorr-Cetina), whose views he opposes by reporting how firmly his physicist interlocutors seem to be committed to the reality of the protons that they accelerate.

In chapter 6, Giere's focus changes. He outlines a decision model for scientists, drawing on well-known psychological studies that question the view that scientists are Bayesians. Giere's alternative is to claim that scientists are satisficers, and he offers an account of how the practice of testing can be interpreted if we view scientists as facing a satisficing decision problem. The two final chapters are intended to explore this idea further. Giere uses his own "ethnographic" findings, based on interviews with research scientists, to analyze work in nuclear physics and in plate tectonics from his own perspective on scientific decision making.

Does all this amount to a cognitive approach to the philosophy of science? Hardly a fully-articulated theory—but perhaps, as I shall suggest below, a program for future development. For much of the book, the appeals to the literature of cognitive science are quite disconnected from the philosophical discussions. Periodically, readers are offered expositions of various experimental results whose relevance to the philosophical problems is never made clear. Most importantly, Giere has no overall picture of the cognitive subject sufficiently detailed to inform his ideas about theoretical knowledge and about decision making.[1] As we shall see, the absence of any such account leads to crucial vagueness in Giere's own treatment of scientific decision.

In my judgment, the most successful part of the book contains Giere's exposition of the semantic conception of theories. Here he uses an illuminating example from textbook physics (the linear harmonic oscillator) to reveal how a small part of classical mechanics can be represented from his favored perspective on theories. Moreover, he is able to show how the conception can do useful work in distinguishing various kinds of realist or empiricist commitments. At this point, however, the picture blurs. Giere's own efforts to advocate constructive realism seem to me to be flawed in important ways.

---

[1] There is a useful contrast here with philosophers who do employ a general perspective on human cognition to reformulate philosophical questions. See, for example, Alvin Goldman's *Epistemology and Cognition* (Cambridge: Harvard, 1986).

According to Giere, use of the semantic conception of theories frees the realist from all kinds of traditional problems about truth and reference. Instead of worrying about correspondence between linguistic (or mental) representations and the world, we can attend to the less troublesome similarity relations between abstract objects (models) and reality. But what is missing in Giere's account (ironically enough given his devotion to the "cognitive approach") is a conception of how the scientist represents either models or reality. Once we put the cognitive agent back into the picture, it is easy to see that all the familiar concerns about correspondence truth recur in wondering how a scientist's representation of a similarity relation between a model and part of reality can correspond to the actual relation. Giere seems to inherit traditional problems for scientific realists and, as a bonus, the difficulty of reference to abstract entities.

Let us now ask how Giere's conversations at the IUCF might support his realism. The strategy seems to be to show that we can only make sense of the scientists whom Giere studied by invoking "entities with roughly the properties physicists themselves ascribe to protons and neutrons" (112). Now, it is not difficult for Giere to amass evidence that his subjects *believe* that they are accelerating protons —and if Giere's thesis were simply the sociological view that scientists believe in entities beyond the range of standard observation, he would have little difficulty.[2] But Giere's attempts to go further seem doomed. Why should we not explain the positioning of the neutron detectors at the IUCF in terms of the physicists' beliefs about what is going on? According to Giere, appealing to beliefs rather than to the facts about neutron trajectories would be "cumbersome" and "substantially weaker" (127). This rejoinder misses the obvious point that where the physicists put the neutron detectors depends on what they *believe* about the situation and what they *want* to achieve, as we can readily appreciate by supposing that they are unaware of what is actually occurring in nature. The very style of explanation that Giere disavows is the appropriate form for explaining behavior. *Explaining Science* does contain some passages that seek to go further, apparently posing the problem of explaining the fact that the physicists' behavior *works* (125). These passages are too brief to support

---

[2] The occasional passages in which Giere suggests that this is what he is about are outweighed by those in which he advances stronger claims That is surely as well. Michael Devitt describes the view that scientists sometimes succeed in referring to mind-independent entities as "anti-realism with a fig-leaf"—*Realism and Truth* (Princeton: University Press, 1984), p. 15 When "realism" is reduced further to the claim that scientists sometimes believe that they are referring to mind-independent entities, the fig-leaf would seem to be transparent.

Giere's conclusions, however, since, as realists know all too well, arguments that appeal to the apparent success of science raise all sorts of complications that Giere does not consider.

If the relevance of cognitive science to Giere's realism is unclear, its *potential* significance for the treatment of decision making is obvious. But the appeal to psychological findings about failures in Bayesian reasoning needs to be handled carefully: although I agree with Giere that investigation of human cognitive practice is important (for 'ought' does imply 'can'), his treatment often seems not to focus on the disclosure of capacities and limitations but on praising the cognitive status quo. But although he appeals to psychological findings to dismiss Bayesianism, he supplies no cognitive picture for his own claim that scientists are satisficers. Distinctions here are critical, for satisficing behavior can be mimicked by Bayesians who maximize subject to constraint, Bayesian maximizing behavior can be imitated by satisficers who have an appropriate satisfaction criterion, and, in the testing situation, Bayesians and satisficers might behave just like old-fashioned hypothetico-deductivists. Indeed, so underdescribed is Giere's account of scientific decision making and so schematic his depiction of the testing situation, that I found it impossible to differentiate his own account from several familiar rivals.

The ethnographic enterprise of the final chapters seems to me to be admirable, for philosophers of science should learn more about how scientists actually behave. As a project for testing Giere's views of scientific decision making, however, it could not succeed, for, given their vagueness, those views are too loosely connected with the available data. Furthermore, the evidence from conversations with nuclear physicists and geophysicists is, though often fascinating, problematic. How much can we learn about the practice of science from interviews with a small sample of scientists? Surely more than from sitting in our armchairs, but equally surely less than from detailed studies of the behavior of all the participants in a scientific field over a period in which there is significant change.[3]

Explaining Science is a book of promises, and the promises are often unredeemed. Yet, despite my criticisms of it, I hope that it will be influential. Philosophy of science needs the cognitive approach, not because psychology has been abandoned in recent philosophy of science but because a crude version of empiricist psychology has been taken for granted. A rapprochement between philosophy and sociology is also overdue, and it is to Giere's credit that he has been able to look beyond the apparently untutored but provocative philosophy in recent sociology of science and to see the value of sociologists' depictions of scientific practice.[4] In short, this book should be read for its suggestions, for the outlining of a far more complex enterprise than Giere could have completed in the space he allowed himself. Giere's questions are, I think, far more valuable than his answers or his arguments.

---

[3] Lest it be thought that such studies are impossible, let me note that one has been carried out. David Hull's investigation of cladistic and phenetic systematics, reported in his *Science as a Process* (Chicago: University Press, 1988). Hull's study faces the potential criticism that his "field subjects" are hardly paradigm scientists—they are involved with questions that are frequently philosophical—but it is methodologically much more sophisticated than the research undertaken by Giere.

PHILIP KITCHER

University of California/San Diego

In Giere's defense, it might be urged that the laboratory studies undertaken by some sociologists of science have the same anecdotal quality There are serious methodological issues concerning the evidential force of such studies, which deserve discussion by both philosophers and sociologists

[4] It does seem to me, however, that Giere's accounts of the writings of sociologists of science fail to do the careful analytic work that a sympathetic philosophical reader ought to provide In particular, his discussions of "contingency" and "negotiation" appear quite inadequate.

NOTES AND NEWS

The Austrian Ludwig Wittgenstein Society announces the Fifteenth International Wittgenstein Symposium in Kirchberg/Wechsel, Austria, August 18–25. The main theme will be "Applied Ethics and Its Foundations." The deadline for receipt of abstracts is May 31. Those planning to attend the conference should contact A. Hübner, Austrian Ludwig Wittgenstein Society, A-2280 Kirchberg/Wechsel, Markt 2, Austria.

The London School of Economics and Political Science announces the creation of a Centre for the Philosophy of the Natural and Social Sciences from October 1990. The Centre's general aim is to promote research and excellence in its field. The Centre will be run by an interdisciplinary Steering Committee comprising members of the Department of Philosophy, Logic & Scientific Method together with other LSE academics. The first director is John Worrall.

Professor and Mrs. Colin Turbayne have established a Berkeley Prize Essay competition in conjunction with the Philosophy Department at the University of Rochester. The next deadline for submitting papers is November 1. Essays should be new and unpublished and should be written in English and not exceed 5,000 words in length. Submissions will be judged by members of a review board selected by the Philosophy Dept. at the Univ. of Rochester. Submissions should be sent to· Prof. Ralf Meerbote, Chair, Phil. Dept., Univ. of Rochester, Rochester, NY 14627.

The Hume Society is pleased to announce a call for papers for its Nineteenth Hume Conference taking place at the University of Nantes, June 29–July 3, 1992. Papers may be on any aspect of Hume's writings, although the conference directors particularly encourage submissions on three themes. "Hume and French Philosophy, Past and Present," "Hume's Aesthetics," "19th and early 20th-century Interpretations of Hume." Papers should be no more than thirty minutes reading length with self-references eliminated for blind refereeing. They may be in English or French, but all papers must be accompanied by an abstract in English. Deadline for submissions is October 1, 1991. Send three copies of papers and abstracts to: Prof. Dorothy Coleman, Secretary of the Hume Society, Phil. Dept., College of William and Mary, Williamsburg, VA 23185.

Penn State Press announces a new series in Literature and Philosophy. The series will publish books in a wide range of subjects in philosophy and literature, including studies of the social and historical issues that relate these two fields. Submissions, proposals, and inquiries may be directed to A. J. Cascardi, Comparative Literature Dept., Univ. of California, Berkeley, CA 94720.

The Department of Philosophy at Bowling Green State University is pleased to announce the appointments of Edward F. McClennen as Ohio Regents Eminent Scholar in Moral and Social Philosophy, and Loren E. Lomasky as Professor.

The Department of Philosophy at Johns Hopkins University is pleased to announce the appointment of Richard Bett as Assistant Professor beginning July 1, following a term as Visiting Assistant Professor in the spring.

The Department of Philosophy at Long Island University/Brooklyn is pleased to announce the appointment of Kristana Arp as Assistant Professor.

The Department of Philosophy at the University of California/Irvine is pleased to announce the appointments of Alan Nelson as Associate Professor with tenure, and Christia Mercer as Assistant Professor.

The Department of Philosophy at the University of California/Davis is pleased to announce the appointments of Paul Teller as Professor, and Michael Dietrich as Acting Assistant Professor.

The Department of Philosophy at University of California/Riverside is pleased to announce the appointment of Georgia Warnke as Professor as of July. The Department also announces the first annual Riverside Colloquium, May 10–11, on ethics of virtue. For more information, you may write to John Martin Fischer, Phil. Dept., Univ. of Caifornia, Riverside, CA 92521.

First Volume in a New Series —
The Collected Writings of Rousseau

ROUSSEAU, JUDGE OF JEAN-JACQUES

DIALOGUES

Edited by Roger D. Masters and Christopher Kelly

Despite the fame of Jean-Jacques Rousseau as a philosopher, writer, and political theorist, there is no single English-language edition of his collected writings. Now, for the first time, his work will be combined in an ambitious, multivolume series. Organized by major themes of his literary career, this invaluable collection will include many texts that have never appeared in English or are newly translated for the first time in 200 years. The series premiers with an original translation of *Dialogues*, in which Rousseau sought to justify his entire career. *Series brochure available.* $40.00

New England
University Press of New England • Hanover, NH 03755 • 800-421-1561

Noûs

Prof. Hector-Neri Castañeda, *editor*
Prof. Earl Conee, *associate editor*

VOLUME 25, NO. 1 **MARCH 1991**

Articles: J. Bigelow/***A Modal Theory of Time;*** S. Davies/***The Ontology of Musical Works and the Authenticity of their Performances;*** G. Landini & T. Foster/***The Persistence of Counterexample: Re-examining the Debate Over Leibniz Law;*** D.M. Armstrong & A Heathcote/***Causes and Laws;*** W. Dembski/***Randomness by Design.***

Critical Reviews: L Aqvist ***by*** J. Tomberlin, E. Pincoffs ***by*** D. Wong; E. LePore & B. McLaughlin ***by*** W. Sinnot-Armstrong; A. Rosenberg ***by*** F Gifford; T. Penner ***by*** G. Fine, R. Meyer ***by*** P Moser; A Donagan ***by*** J. Montmarquet; E. Wolgast ***by*** E Byrne; H. Gadamer ***by*** J. Zuñiga, J Tomberlin ***by*** T Horgan, Annas & Barnes ***by*** E De Olaso.

VOLUME 25, NO. 2 **SPECIAL APA ISSUE** **APRIL 1991**

Symposium Papers and Abstracts of Invited Papers by: J Wallace; B. Herman, P. Railton, D. Schweickart; C Gould; T. McCarthy, C Card; L. Outlaw; O Flanagan; M. Friedman, R. Mohr; D. Kolb, C. Wright; M. Kaplan; W Lycan; D. Sanford; D Nute, V. McGee.

Critical Reviews: N. Cocchiarella ***by*** D. Bonevac, J. Elster ***by*** M Fisk; M. Bratman ***by*** H.J. McCann; G B Madison ***by*** J. DeCew; T.V. Morris ***by*** B. Miller; L. Holscher ***by*** M Strasser, G Schlesinger ***by*** P. Quinn; H. Brown ***by*** J. Fetzer; K Hanson ***by*** R Lemos.

Subscribe to:
Noûs % Secretary
Submissions to:
Noûs % Editor

Dept. of Philosophy
Indiana University
Sycamore 126
Bloomington, IN 47405

1991 Subscriptions:
Institutions. $60 00
Individuals: $30.00
Single Issues· $12.00

Oxford

The Language of Morals
R.M. HARE
"It is a perceptive contribution toward the solution of many fundamental problems of ethics." —*Philosophical Review*. "It is indeed the accuracy, the thoroughness of this analysis that makes *The Language of Morals* by far the best introduction to moral philosophy to appear for many years."
—*New Statesman*
1991 208 pp. paper $14.95

Contemporary Political Philosophy
An Introduction
WILL KYMLICKA, *University of Toronto*
Offering a critical appraisal of utilitarianism, liberal equality, libertarianism, Marxism, communitarianism, and feminism, Kymlicka argues that these philosophies can best be viewed as various interpretations of the principle that all people should be treated as equals.
1991 336 pp. paper $13.95 cloth $39.95

Philosophical Applications of Free Logic
Edited by J. Karel Lambert, *University of California, Irvine*
In this collection, Karel Lambert, one of the pioneers in, and the most prominent exponent of, free logic, brings together a variety of published essays bearing on the application of free logic to philosophical topics ranging from set theory and logic to metaphysics and the philosophy of religion.
1991 288 pp. $39.95

Apologie des Zufallingen
(In Defense of the Accidental)
ODO MARQUARD, *University of Griessen, Germany*
Translated by Robert M. Wallace
The seven essays collected in this book address the history of modern ideas and contemporary cultural issues, ranging from the discourse of Marquard's acceptance of the Sigmund Freud prize to the notion of world history to a mediation on chance and luck as essential aspects of the human condition.
April 1991 144 pp. $27.50

Frege and Other Philosophers
MICHAEL DUMMETT, *Oxford University*
This book comprises all of Dummett's published and previously unpublished essays on Gottlob Frege, with the exception of those included in his *Truth and Other Enigmas*. Several of the essays illustrate his contention that Frege's work remains the best starting point for the investigation of problems concerning truth, meaning, and thought and language, which are still among the most contentious issues in modern analytical philosophy.
1991 344 pp. $39.95

New in paperback!
Religious Convictions and Political Choice
KENT GREENAWALT, *Columbia University*
"A searching, well argued, and richly suggestive treatment of the ethics of citizenship in a liberal democracy.... The book contains much that is of value in basic social-political philosophy. It is also lucid, forthright, and concretely written." —*Ethics*
1987 (paper 1991) 256 pp. paper $13.95 cloth $34.50

Prices are subject to change and apply only in the U.S.
To order, send check or money order to
Humanities Marketing Department, Dept. MG
To order by phone using major credit cards please call
(212) 679-7300, ext. 7106

Oxford University Press
200 Madison Avenue • New York, NY 10016

Classics in Philosophy

GREEK PHILOSOPHY
2nd Edition
Reginald E. Allen
ISBN. 0-02-900660-0 $12.95 paper

POLITICAL IDEAS OF ST. THOMAS AQUINAS
St. Thomas Aquinas
ISBN 0-02-840380-0 $10.95 paper

EIGHTEENTH CENTURY PHILOSOPHY
Lewis W. Beck, Editor
ISBN· 0-02-902100-6 $12.95 paper

A MODERN INTRODUCTION TO PHILOSOPHY, 3rd Edition
Paul Edwards & Arthur Pap, Editors
ISBN: 0-02-909200-0 $27.95

A SHORT HISTORY OF CHINESE PHILOSOPHY
Fung Yu-Lan
ISBN 0-02-910980-9 $14.95 paper

NINETEENTH-CENTURY PHILOSOPHY
Patrick Gardiner, Editor
ISBN 0-02-911220-6 $12.95 paper

INTRODUCTORY READINGS IN AESTHETICS
John Hospers
ISBN 0-02-915260-7 $14.95 paper

MORAL AND POLITICAL PHILOSOPHY
David Hume
ISBN 0-02-846170-3 $12.95 paper

DIALOGUES CONCERNING NATURAL RELIGION
David Hume
ISBN 0-02-846180-0 $10.95 paper

ESSAYS IN PRAGMATISM
William James
ISBN 0-02-847140-7 $10.95 paper

CRITIQUE OF JUDGMENT
Immanuel Kant
ISBN 0-02-847500-3 $12.95 paper

JOHN STUART MILL'S PHILOSOPHY OF SCIENTIFIC METHOD
John Stuart Mill
ISBN 0-02-849250-1 $12.95 paper

PHILOSOPHY AND TECHNOLOGY
Carl Mitcham and Robert Mackey, Editors
ISBN· 0-02-921430-0 $14.95 paper

THE SPIRIT OF THE LAWS
Baron de Montesquieu
ISBN 0-02-849270-6 $13.95 paper

NEWTON'S PHILOSOPHY OF NATURE
Isaac Newton
ISBN 0-02-849700-7 $12.95 paper

A CRITICAL HISTORY OF WESTERN PHILOSOPHY
D.J. O'Connor
ISBN: 0-02-923840-4 $14.95 paper

PHILOSOPHY OF THE SIXTEENTH AND SEVENTEENTH CENTURIES
Richard H. Popkin
ISBN· 0-02-925490-6 $12.95 paper

GREEK AND ROMAN PHILOSOPHY AFTER ARISTOTLE
Jason L. Saunders
ISBN. 0-02-927730-2 $14.95 paper

ETHICS
Benedict de Spinoza
ISBN 0-02-852650-3 $10.95 paper

TWENTIETH CENTURY PHILOSOPHY
The Analytic Tradition
Morris Weitz
ISBN 0-02-934990-7 $12.95 paper

MEDIEVAL PHILOSOPHY
John F. Wippel & Allan B. Wolter
ISBN 0-02-935650-4 $12.95 paper

Attention College Professors Examination copies will be sent on a 30-day approval basis only Simply place your request on university stationery, indicating book requested, course title, projected enrollment, and date your next decision is due. When ordering examination copies of paperbacks, enclose a check for $2.00 per copy to cover postage and handling
Mail to The Free Press, Attn Lisa Weinberg, 866 Third Avenue, NY, NY 10022

Fp **THE FREE PRESS**
A Division of Macmillan, Inc
866 Third Avenue, New York, New York 10022

Detail from The Plum by Edouard Manet Courtesy of the National Gallery of Art Washington, D C

LOVE AND ITS
PLACE IN NATURE
JONATHAN LEAR

A PHILOSOPHICAL INTERPRETATION
OF FREUDIAN PSYCHOANALYSIS

"A wonderful, brilliant, and wise book."

— Leonard Shengold, M D
author of *Soul Murder*

"Jonathan Lear makes one understand how psychoanalysis works not only on the therapist's couch but also as a condition of being alive *Love and Its Place in Nature* not only offers a form of spiritual nutriment for the self, it also defines that self with a clear profundity that few readers will have encountered before."

— Christopher Lehmann-Haupt
The New York Times

FARRAR · STRAUS · GIROUX
Available at bookstores everywhere or call 1-800-631-8571

New From Cambridge University Press

Richard Rorty's Philosophical Papers in Two Volumes:

These two volumes collect the recent philosophical papers of Richard Rorty, considered by many to be the most interesting philosopher writing today.

Objectivity, Relativism, and Truth

Philosophical Papers, Volume I

Richard Rorty

The volume is divided into three parts. In Part I Rorty sets out an antirepresentationalist account of the relation between natural science and the rest of culture, arguing that science is a matter not of discovering a "true," "objective" reality independent of mind and language but rather of acquiring habits of action for coping with reality. In Part II Rorty explores the work of the contemporary philosopher who he thinks offers the best account of objectivity and truth—Donald Davidson. Part III extends these pragmatist views into politics by discussing the needs of a democratic society.

35369-6 Hardcover $39.50
35877-9 Paper $12.95

Essays on Heidegger and Others

Philosophical Papers, Volume 2

Richard Rorty

The second volume discusses a group of thinkers who have broken free from the traditional representationalist doctrines of philosophy. Most of the essays deal with Heidegger and Derrida, about whom Rorty is critical when he feels they are becoming quasi-metaphysical. The volume also discusses Freud and moral deliberation, as well as the social theories and political attitudes of Habermas, Unger, Castoriadis, and Foucault.

35370-X Hardcover $39.50
35878-7 Paper $12.95

Two-volume set
40476-2 Hardcover $70.00
40915-2 Paper $22 00

Available in bookstores or write:

CAMBRIDGE UNIVERSITY PRESS
40 West 20th Street
New York, NY 10011.
Call toll-free 800-872-7423.
MasterCard/VISA accepted.
Prices subject to change.

NEW FROM CHICAGO

Crossings

Nietzsche and the Space of Tragedy

John Sallis

Boldly contesting recent scholarship, Sallis argues that *The Birth of Tragedy* is a rethinking of art at the limit of metaphysics. His close reading focuses on the complexity of the Apollonian/Dionysian dyad and on the crossing of these basic art impulses in tragedy.

Paper $11.95 168 pages
Library cloth edition $24.95

The Rhetoric of Morality and Philosophy

Plato's *Gorgias* and *Phaedrus*

Seth Benardete

Distinguished classicist Seth Benardete here interprets and, for the first time, pairs two important Platonic dialogues, the *Gorgias* and the *Phaedrus*. Taken together, Benardete suggests, these dialogues construct a complete psychology.

Cloth $29.95 216 pages

Science and Relativism

Some Key Controversies in the Philosophy of Science

Larry Laudan

Laudan's fictional dialogue among a realist, a positivist, a pragmatist, and a relativist is a trenchant critique of cognitive relativism and a thorough introduction to major issues in the philosophy of knowledge.

Paper $12.95 194 pages
Library cloth edition $32.00
Science and Its Conceptual Foundations series

Now in Paper—

Of Spirit

Heidegger and the Question

Jacques Derrida

Translated by Geoffrey Bennington and Rachel Bowlby

Derrida exposes the *political* meaning of Heidegger's avoidance and invocation of the word "spirit" and his preference, before 1933, for the term "being." He discusses in particular how "spirit" is linked to the ideology of the Nazis.

"Let there be no mistake: this is not merely a brilliant book on Heidegger, it is thinking in the grand style."—David Farrell Krell, *Research in Phenomenology*

Paper $9.95 152 pages

The University of Chicago Press

5801 South Ellis
Chicago, IL 60637

THE JOURNAL OF PHILOSOPHY

Philosophy

VOLUME LXXXVIII, NUMBER 4
APRIL 1991

page
169 *How Euclidean Geometry Has Misled Metaphysics* Graham Nerlich
190 *Agency and Morality* Richard Brook

BOOK REVIEWS
213 Paul Grice: *Studies in the Way of Words* Robert J. Fogelin
219 Paul Feyerabend: *Against Method & Farewell to Reason* Ian Hacking

223 NEW BOOKS: ANTHOLOGIES

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1991 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy*, 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NO. 4, APRIL 1991

HOW EUCLIDEAN GEOMETRY HAS MISLED METAPHYSICS*

EUCLID'S geometry is so brilliant that it has always dazzled us. It shone for centuries so brightly as to blind us to any model for an established branch of knowledge but the deductive system based on a few self-evident axioms. Thus, it misled metaphysics. But that is not my theme. I want to show how it led Leibniz to some mistaken arguments in the central letters of the Leibniz-Clarke correspondence. The letters are about quite special issues in the metaphysics of space. But they have dominated philosophy since then as our models of how epistemological arguments should decide metaphysical questions. My theme is to reveal the arguments as mistakes.

Space sets ontology one of its most acute, searching, and elegant problems. To change my visual metaphor, the reason why we have not been able to see further into this subject is because a giant has been standing on our shoulders. He is Leibniz. Although the arguments in his letters are fallacious, surely he is a giant. What can make those fallacies clear to us was not available to him nor to his great opponent Newton. But first: What is the problem space sets for ontology?

I. INTRODUCTION

On the face of it, space has disconcerting properties. The list below is loosely phrased, and neither exclusive nor exhaustive. But it shows us what the problem is—on the face of things, at least.

---

* I have read versions of this paper to a number of meetings in Australia, Canada, and the United States. It owes much to the stimulating discussions it received I am especially grateful to my colleagues at the University of California/San Diego, where I wrote the first version of it while on leave there.

0022-362X/91/8804/169–89

(i) Space is a concrete *particular,* being neither general nor abstract. It has parts not instances. There is only one space (we ordinarily think) within which everything is related to everything else.

(ii) Space stands in *concrete* relations to concrete things. I *move through* it, *see across* it to trees and cars at various distances, it *contains* me and all these concrete things. These are relations which, typically, only concrete things have to each other. Abstract things can play no part in them.

(iii) Space is *physical*—that is, space plays a prominent role in physics. Physicists are always tracing paths and trajectories, measuring lengths or constructing theoretically sophisticated, elaborate, and expensive standards of length invariance. It also figures in their theories

(iv) Space is *not material.* As classically conceived, it has no mass, momentum, or energy. Our usual nomenclature for views of the nature of space is unhelpful. Newton denied that space is a substance just because it is not material in this sense—"it does not support characteristic affections of that sort that denominate substance, namely actions such as are . . . motions in a body."[1] That seems right. If so, then the view that it acts, is material, or has mass or energy surely ought to be called substantivalism. That is the right way to describe space-time as we find it in General Relativity.

(v) Space is *not perceptible.* We never see space or its parts; we neither hear them, smell them, nor touch them. Yet we can see across it. We can see at a glance that just two objects are now visible though we "can't tell them apart." Suppose that they are in fact qualitatively identical. That surely means that we can see immediately that the two are in different places even though we see only the things and not the places.

The acute problem for ontology lies in the sharp clash between the first three properties on the list and the last two. The first three properties seem to tell us that space is something; the last two that it is nothing. This problem about space in our ontology is unresolved. *Realism* is my name for the view that space is a thing, earns a place in our ontology, and has (something like) properties mentioned in the list. I distinguish realism from *substantivalism.*

The precision and power of geometry make the space problem elegant because the arguing can be so neat, clear, and determinate. The curious relations of space to perception (feature (v)) demand that we frame with particular care and rigor the epistemological principles on which we either reject or accept an ontic commitment to space. So the problem is searching, too.

---

[1] For a corrected translation of a part of the *Unpublished Scientific Papers of Isaac Newton,* A. Rupert Hall & Marie Boas, eds. (New York: Cambridge, 1962), see Howard Stein, "On Metaphysics and Method in Newton" (forthcoming).

II. DOUBLING IN SIZE

Let us begin with a familiar and graphic challenge. It snugly fits the mould of Leibniz's arguments, though we owe the first form of it rather to Henri Poincaré. Both the style and the mistake are the same as in Leibniz's letters. Every philosophy student surely hears this argument at an early stage of his education. It is one of our best-known intuition pumps. It will be useful to start with it.

Suppose everything were to double in size overnight. Nothing would betray the change. Round pegs would fit snugly into their old round holes and other holes squarely resist their entry as before. Everything would be related just as it was to everything else. The difference would really be no difference, the supposed change no change at all. Nocturnal expansion is absurd.

In more detail, the argument has to run like this. At first glance, space looks a useful posit: the size of a thing is the quantity of space it fills, its shape is the disposition of the thing's parts in space, the spatial relations of one thing to others is fixed by their places in space and the lengths and directions of the spatial paths that join one place to others. The spatial relations between things look as if they are all mediated by relations of things to space.

The doubling argument seeks to cut this tie. It is a kind of reductio argument in which the absurdity is not a contradiction but an idleness. It is a thought experiment in which the realist gets his space but finds that he can do nothing with it. It shows how to change the thing-space spatial relations of every thing yet retain all thing-thing spatial relations. Conversely, the thing-thing spatial relations are shown not to determine any thing-space spatial relations. The relationist's gift of space is revealed as ironic—it gives nothing at all. There are no thing-space relations. So it is needless to posit an entity that exists only to found them. Space is a chimera.

All this rests on some kind of indiscernibility or inconsequentiality principle. It might be this one:

IP: "Inconsequential properties" are not properties.

Here 'inconsequential property' means something like "property which can change without any accompanying change in specifiably privileged properties." For the case in hand, we have two ways of saying what is privileged. Observable properties are privileged; more specifically (and less naively), thing-thing spatial relations are privileged. I shall stick to the latter account which will do for Leibniz himself. So let us distinguish thing-thing spatial relations from thing-space spatial relations. Thus, we can state the detachment thesis:

Detachment thesis: thing-thing spatial relations are logically independent of thing-space spatial relations.

This will put us in a position to deny the realism of the introduction. All thing-space spatial relations are idle; so the relatum of these relations is idle, too. Space is no part of our ontology. Only the concrete things that stand in spatial relations together with those relations themselves are real; space is an unwanted absurdity.

The detachment thesis lies at the foundations of a prominent reductive theory of metaphysics: spatial relationism. Spatial relationism can fairly promise to reduce space to thing-thing spatial relations as long as these relations can be detached from thing-space spatial relations. The detachment thesis clears relationism of the otherwise plausible charge that it is circular. The detachment thesis is no help in showing how the relationist can complete his programme. It simply legitimates his title to use thing-thing spatial relations in the reduction. It offers an escape from circularity. It is not obvious how to escape circularity without it.

III. OBJECTIONS TO THE DOUBLING ARGUMENT

Perhaps the reason why Leibniz never used the doubling argument is that he read the first of Galileo's *Two New Sciences.* Galileo points out that doubling a thing in all its lengths (scaling it up) will quadruple its cross-section areas and multiply its volume and thus its mass by eight. The load-bearing factors of the object are essentially functions of area, but the loads they must bear for the thing to be stable are functions of its mass. So scaled-up models of strong small things, like ants, may be unstable, as producers of gigantism horror movies often find. The bigger they are the harder they stand—or fall. What, too, about the speed of light, and other constants—Planck's for one?

Poincaré thought that he could avoid this objection and it is evident that Hans Reichenbach, Adolf Grünbaum, and many others thought so, too. This was challenged by George Schlesinger.[2] I shall not try to umpire that debate, though it surfaces again briefly in section VI, objection 2. My main objection, in the next paragraph, is simpler and independent of this. I think it is more incisive.

The claim that doubling has no consequences rests on a quite precise and unique assumption about thing-space relations. Things must be doubled *in Euclidean space.* Among the geometries of constant curvature, only Euclid's geometry allows similar figures of different sizes. If everything doubled in size overnight in a non-Euclidean space, everything would change in shape. Thing-thing spatial relations would change with their changing thing-space relations. Further, in non-Euclidean spaces, doubling in size would differ observably from tripling in size in non-Euclidean spaces—it induces weaker changes in shape than tripling in size. The geometrical facts are familiar enough. But their significance for metaphysics has gone unnoticed. Before I turn to that, I shall illustrate the geometrical facts as simply as I can in the next, starred paragraphs. The reader who does not need the illustration can skip them.

---

[2] "What Does the Denial of Absolute Space Mean?" *Australasian Journal of Philosophy,* XLV (1967): 44–60 And cf. Adolf Grünbaum, "The Denial of Absolute Space and the Hypothesis of Universal Nocturnal Expansion: A Rejoinder to George Schlesinger," *Australasian Journal of Philosophy,* XLV (1967): 61–91.

* It is easy to see what doubling in size comes to in two-dimensional spaces (surfaces) that do not have the character of the Euclidean plane. The consequences generalize fairly intuitively to their counterparts in three dimensions. Let us begin with the surface of a sphere. A triangle in that surface will be any three-sided figure composed of sides that are geodesics of the sphere, that is, great circles of it. If the triangle is small relative to the radius of curvature of the space, then the sum of its three angles will be close to a straight angle. But a big triangle will have no such shape. Imagine, as one side of a large triangle, an arc of a great circle just less than half the circumference long. At its end points, construct two more great circles to meet at a point polar to this first side of the triangle. The angles at each of the vertices on the equator will be a right angle (half a straight angle), so together they will sum to a straight angle. The angle at the pole will be close to a straight angle, too, so that the whole angle sum will be slightly less than 360°—nearly a complete revolution. That is a gross (but not the grossest) departure from Euclidicity. This defect (as it is called) is a function of the ratio between the area of the triangle and the radius of the curvature of the space So, if a triangle on the sphere is doubled in size overnight, it changes shape. Not only is its shape different; it now has a shape *it could not have had* yesterday.

* Spaces of positive constant curvature, of which the spherical surface is a two-dimensional example, are finite spaces. To double everything seems to mean doubling the size of configurations of objects (e.g., stellar triangles) too. But then there might not be enough space for them to double in. If stars are spread more or less uniformly all through a finite space, the experiment is not thinkable. And if the configurations are not doubled, the doubling just of star sizes will obviously change thing-thing spatial relations. Some spaces of constant zero curvature are finite, too (the torus, for instance). So, even when the local character of space is everywhere Euclidean and doubling is locally indiscernible, we cannot guarantee that it will be globally indiscernible. The operation cannot be imagined as performable no matter what geometry is in question. We have to know what the space we are dealing with is like.

* We get defects from doubling lengths in negatively curved spaces, too, though here the angle sums get smaller as the lengths increase.

* If triangles change in shape on doubling in these spaces, then so does every other shape. Any polygon can be decomposed into triangles;

any closed curve can be approximated by some polygon. As surfaces change in two dimensions, so do the volumes they bound in three dimensions; and so on up. The result is perfectly general.

* If the space varies in its curvature from point to point, then the consequences of doubling will depend on where the doubling begins and how it spreads. Imagine, first, a surface that varies its curvature from point to point: it will resemble a range of smooth hills and valleys. In some places, the curvature will be positive (hill tops and valley bottoms); in other places (saddle backs or hill sides), the curvature will be negative and, thus, since we suppose the curvature varies smoothly, some places in between will have zero curvature. Now, it will certainly make a difference to thing-thing spatial relations from which center (whether occupied or empty) the doubling occurs. If one point rather than another is the center, then some things will be moved further than others by the expansion and that, in turn, will determine whether they themselves are doubled in regions of positive, negative, or zero curvature. So there will be uncountably many ways of performing the thought experiment and uncountably many distinct thing-thing relations resulting from them. So doubling-in-thought experiments depend on which thing-place relations of spatial occupancy (position, place) are preserved, if any, and how others are altered.

* Perhaps it is worth pointing out that such changes are in general quite grossly observable. How are we to measure the angle sums of large triangles? Here is a crude method. At each angle, use a stretched string to draw a circle of standard radius by anchoring one end of the string at the vertex. Now lay the string along the drawn circle with one end of the string at the point where one side of the triangle cuts the circle and mark along the string where the other side cuts the circle. So we measure the length of the arc of a unit circle that coincides with the rotation through that angle. (The length of the chord would serve equally well provided we note when the angle measured by it is greater than a straight angle.) We can transfer the results of this operation from one vertex to another and add the angles simply by adding the arc lengths successively round the unit circle. The one piece of string will give us both the unit circle and the arc lengths. Now, if we carry out this operation of adding the two distant angles of the large triangle to the nearby one, we shall find that the arcs add to almost a complete revolution (or more for still larger triangles) There is nothing subtle about the defects.

The Leibniz-style Poincaré objection to the realist view of space is fallacious. It is an argument to the detachment thesis—that thing-thing spatial relations are independent of thing-space spatial relations. But the crucial invariance of thing-thing spatial relations requires attachment, that is, precise and unique thing-space spatial relations. Things are to be expanded *through Euclidean space*. Otherwise, it will not even appear to legitimate its claims to inconsequentiality. What it presupposes is inconsistent with what it concludes. Note in particular that the dependence of the outcome of doubling on the nature of the space applies with equal force to the special and actually indiscernible case of doubling in Euclidean space. It is not only when doubling will change shapes that we need to specify in which space we are performing the thought experiment. That the doubling operation in Euclidean space has no thing-thing consequences depends on specific symmetries unique to the structure of that space. These symmetries pick out the similarity transformations from the rest. They do not show that there is no structure and no space. They mean that there are specific structures in a specific space.

The string of words 'A is double the size' is not a sentence; it is not syntactically complete. If you like, it is a chimera. To have a thought, true or false, we have to know what the thing is twice as large *as*. Just so, the "thought experiment" given by the directive 'Double things in size overnight' must be syntactically complete. But now it is not enough to know that things are to be double the size they were. We need to know in (through) what they are to be doubled. In each case, the chimera is a monster—ill-formed as a sentence or a thought. The imperative that directs us to perform the operation is syntactically incoherent. The "thought experiment" is unintelligible. We have to specify the space. (Of course, we shall get general sense out of the idea of doubling things in *some* space, since that is syntactically coherent. But like the idea of painting a thing some color, it tells us little about how things will look when the job is done.)

Further, we have to double things in space (expand them through it) to hope to support a detachment thesis: else it will no longer be a thought experiment that detaches thing-thing from thing-space spatial relations. The relationist has to begin by giving the realist a space. His problem is how to get this gift back again.

IV. LEIBNIZ AND THE DETACHMENT THESIS

Leibniz brought several arguments against Clarke to "confute the fancy" of those who posit the chimera of space. All of them founder on the same rock and by much the same style of navigation as we have seen. The arguments need symmetries peculiar to Euclidean space. The appeal fails without them. (Leibniz ties his arguments to a theological dispute which is beside the point of this paper. Perhaps theological issues warp any straight relevance to modern issues. In that case, what I have to say will concern interpretations of his arguments in the more recent philosophy of science.)

Let us begin with the most important of these, the relativity of uniform motion in Newtonian mechanics.[3] How does this engage the inconsequentiality principle? To every body in the universe, whatever its state of rest or motion, add in thought the same velocity vector—the same speed in the same direction. Then everything will change its thing-space spatial relations; everything will be moving through space. Leibniz claims that this will make no difference to thing-thing spatial relations.

Newton's mechanics entails Euclidean space. It presupposes that we *can* add the same velocity vector to every thing. We can do that only if each added vector is parallel to every other. But the existence of parallels is peculiar to Euclidean geometry.

Consider a cloud of small particles that move independently (do not exert significant forces on each other). It is dust. Then, like its component particles, the cloud itself can move at uniform velocity through Euclidean space, free from the action of any force. It will change its thing-space spatial relations (position) continuously but the relative position vectors among the particles in it will not change; they are the vectors that relate particles in the cloud to each other at any time by their distances and directions. So the cloud does not change its shape as it moves. In motion, it is just as if the cloud were at rest. That, in turn, is possible because every particle in the cloud can have the same velocity vector. But the idea of the same velocity vector makes sense only if the velocity vector on any one particle is parallel to that on any other. Mechanics allows a symmetry among the relative uniform motions of point particles. That is Galilean relativity. But we need Euclidean geometry to extend this to moving volumes. In other geometries of constant curvature, there are no parallels. This may be because all geodesics intersect somewhere (positive curvature). It may be because, given any geodesic G and a point P not on G, there will be infinitely many other geodesics through P, coplanar with G and P, and which do not intersect G anywhere. That is the case when the curvature is negative. Nevertheless, there will be no geodesic in the plane and through the point that is everywhere the same distance from G. It is often said of spaces of constant negative curvature that they have many parallels. This is not true in the sense needed for the thought experiment if it is to yield the detachment thesis.[4]

---

[3] Leibniz's Fourth Paper, §13; Fifth Paper, *The Leibniz-Clarke Correspondence*, H. G. Alexander, ed. (New York: Barnes & Noble, 1956), pp. 38, 73–5.

[4] For a more detailed account of what the motion of dust clouds and elastic solids might be in spaces of constant nonzero curvature, see my "What Can Geometry Explain?" *British Journal of the Philosophy of Science*, xxx (1979): 69–83.

If a dust cloud moving in non-Euclidean space changes shape, that is not an action of space on things; no more so than the constancy of shape of a moving dust cloud in Euclidean space is an action of space. If the motion of each particle is uniform (as I have been imagining), each will move along a geodesic. That divergence of geodesics in non-Euclidean space is not something that the space does to things that move along them. Space is not a cause, and the explanation of the changing shape of the cloud in motion is geometrical not causal.

Leibniz claims that a uniform motion of simply everything would be without consequences. But this depends on the objects being attached to (being moved through) a quite precise and unique kind of space. Our thought experiment does not show that we can freely detach thing-thing spatial relations from thing-space spatial relations. The relations depend on each other. The argument to the thesis is fallacious; it presupposes (tacitly entails) the attachment it attempts to deny. Here again, quite clearly, the attachment is presupposed not just in the case of non-Euclidean spaces where uniform motion is well-defined and observable. The inconsequentiality Leibniz needs is a symmetry of Euclidean space. He needs the symmetry to be concretely there in something real and not chimerical.

This suggests that Leibniz simply assumed that Euclidean symmetries were inescapable. But it is not quite so simple. He seems to have thought he could derive the symmetries just from the indiscernibility of spatial points one from another. His "proof" runs as follows.

> Space is something absolutely uniform; and, without the things placed in it, one point of space does not absolutely differ in any respect whatsoever from another point of space. Now from hence it follows (supposing space to be something in itself, besides the order of bodies among themselves) that 'tis impossible there should be a reason why God, preserving the same situation of bodies among themselves, should have placed them in space after one particular manner, and not otherwise (Fifth paper §5).

The passage does not make it clear whether Leibniz thinks the uniformity of space derives from the indiscernibility of points or vice versa. But we should not accuse him of taking the symmetries for granted simply from Euclid's geometry.

In any event, his conclusion does not follow. Spatial structures richer than topology are not gained by adding to properties of points. Affine and metric structures are properties of the space as a whole and "emergent" relative to properties intrinsic to points. The presence, absence, or variation of curvature is not an intrinsic difference among points. It depends on the structure of the space. Curvature is a limiting property of points, deriving from the (possibly varying) structure of the space round them.

Leibniz also tells us that it could make no difference to thing-thing spatial relations if everything were displaced some arbitrary distance and all in the same direction. As before, the phrase 'all in the same direction' requires parallels. But there are new problems. He tells us that there would be no change to thing-thing spatial relations if things were interchanged East and West.[5] (Whether this is supposed to be the result of a reflection, or a rotation round an axis makes no odds.) These displacements are not to be thought of as processes of moving things to new places. Leibniz's theological metaphors make the point clear. It would make no difference had God created things not where he did, but so that each came into being 100 meters further North or with East and West interchanged. (Put aside proper worries about North, East, and West as the same directions everywhere.) The point is that the displacement would yield differences if the curvature of the space varies from point to point. Indeed, if our thought experiment asks us to imagine perfectly rigid bodies, not even God will be able to perform it. A rigid body cannot be relocated in a region where the curvature is different, since no shape in a region of one curvature is the same as any shape in a region of different curvature. An elastic body might be relocated by distorting it, but a *perfectly rigid* body cannot be distorted. God cannot act a contradiction. Among elastic bodies, we can expect rubber objects to distort more than steel bodies and expect that it would be easier to relocate them. But we can detect and measure these various distortions. We are not dealing here with Reichenbach's universal forces, but with common or garden elasticity, tensile strength, and the like.

The inconsequentialities of Euclidean space are not the product of a confused hypothesis about a nonentity. They are the symmetries of a unique and well-defined entity with a perfectly definite, if bland, structure. Change the structure of that entity and you have the thing-thing relation consequences that Leibniz thinks impossible. We cannot argue as he does with the inconsequentiality thesis as premise. It is false.

V SOME CONCLUSIONS SO FAR

The detachment thesis lacks support. Any version of the reductive relationist program for banishing space from our ontology which depends on Leibnizian arguments for the detachment thesis is circular. The relationist thinks that his hardest job lies in completing the reduction. But that is really the easy part. The relationist uses what seem at first glance to be parts of space (such as the paths across which two things are at a distance or volumes of space in which the matter of things is disposed). If he can show his entitlement to use them without circularity then his residual problem looks soluble. It is the problem how to determine the character of a whole by specifying relations among certain parts of it. It is by no means a contemptibly easy problem. But it is not obviously a metaphysical one either. The vexed question lies elsewhere in the use of the relations for reductive purposes. Can we detach thing-thing spatial relations from thing-space ones quite straightforwardly? The prominence of Leibniz-style arguments in the literature suggests that, on the face of it, we cannot; these arguments aim to show something that needs to be shown. Suppose that it cannot be shown. Then, as ontologists, we should be no less worried by the nature of the parts of space (volumes, paths, points) than by the nature of the whole. Whole and parts equally have the same awkward features listed in section I. The real problem is to show that thing-thing spatial relations do not depend on thing-space relations. Relationism without a detachment thesis to establish its freedom from circularity lacks reductive teeth.

---

[5] Leibniz, Third Paper, §5, Alexander, ed., p. 26

Leibniz appeals to the "indiscernibility principle"; it is dubious. Its converse is highly plausible, however. We might call it the "discernibility principle" and express it as follows:

DP: Consequential properties are real properties.

This is the contrapositive of the converse of IP. We can gloss 'consequential property' after the style of 'inconsequential property'. Suppose we adopt DP. Then we have really a rather plausible argument that position in space and the quantity of space filled or occupied by a material thing are real relations. So some thing-space relations are genuine and not chimerical. In general, change in these relations changes thing-thing spatial relations in consequence. If the latter changes are real, so are the changing properties that induce the consequences. But then the thing that has the properties must be real, too.

Thirdly, and more generally (but more vaguely), Leibniz's conclusion has been our paradigm for a whole range of epistemological claims about the probity and point of concepts in science and outside it. The path for many of these has been the easier because of the apparently brilliant success of their Leibnizian forerunners. Reductive epistemology did not begin with Leibniz, but his incisive attack on Newton wrote a distinctive new page. This, together with the fact that classical physics after Newton seemed to endorse his attack, lent a respectability to reductive empiricism, in one form or other, which it would hardly have gained otherwise.

Theses that something or other is unobservable in principle have proved vulnerable. Most of them were proposed as having a status unhappily like that of synthetic necessary truths, though, like the verification principle, used to deny that there could be any such truths. They are as fragile as allegedly synthetic necessary truths deserve to be. Yet they have motivated various draconian principles by means of which seemingly respectable conceptual structures were run out of town. This vigilante epistemology has seemed judicious largely because Leibniz's dazzling arguments about space supported it. If Leibniz's arguments are our paradigm for the probity of inconsequentiality and reductivism, their failure might become our paradigm for impeaching them.

Thus, the role of epistemology in metaphysics has been regularly misunderstood and overstated by confining ontology to a dependent and sharply limited role in a puritanically conceived theory of knowledge. In this mistaken and destructive idea of things, a fixation on Euclidean geometry played no small part. It is ironical to reflect that it is the very simplicity and symmetry of Euclidean geometry which has misled us. The simplest possible spatial world is confusing in its very simplicity and blandness.

VI FURTHER ASSUMPTIONS AND CONSEQUENCES

Ontologically, Euclidean space is surely just the same kind of thing as non-Euclidean space. Newton's theory demands that we include Euclidean space in our ontology. The metaphysical nature of this commitment is no more disturbing than when our theory commits us to a more readily observable non-Euclidean space. No doubt Euclidean space is bland. But that cannot change the nature of an ontic commitment to space. Non-Euclidean space has all the disconcerting features in our list of section I just as surely as Euclidean space has them. So I make the following assumption:

> A1: Structural symmetries do not affect the metaphysical nature of an entity.

The structural symmetries of the entity may well alter the way it enters into our theories, however. We now see that it is the realist who can postulate spatial symmetries and thus explain why, in Euclidean space, we can never observe the difference between doubling and constant size, between rest and uniform motion, and so on. The realist owns the symmetries, not the relationist. It is realist's realism that tells him (and tells us) why we cannot care about the difference between one of these states and another. In fact, it is the relationist who has no clear, coherent, noncircular account of the relativity of motion. That relativity rests, not on nothing, but on symmetries of a real entity, Euclidean space. *Therefore, the realist does not have to answer the objection that is standardly brought against him.* The objection is that the realist is obliged to make these inconsequential distinctions and must take them seriously, whereas the relationist need not. But, on the contrary, the realist is just the person who need not concern himself with them, since he and he alone has a clear picture of their source and their nature and their lack of consequences. He alone has a coherent account of when they hold and how they may fail. They spring from the character of Euclidean space. How can anyone oblige him to care about what *the nature of his postulate* makes plain to him that he *cannot* care about?

Let us use ***absolutism*** to describe the view that there are some symmetry transcending properties that matter for science or for metaphysics. Then realism does not entail absolutism. Rejecting absolutism takes no step toward relationism. Newton was an absolutist.

Nonetheless, motion is motion-through-space. 'x moves' is good grammar, for we can say whether it is true or false if it is contextually clear what the tacit relatum is. But the deeper syntax of sentences of motion requires that the referent be filled out on demand—the demand of how the sentence is to be understood. Now, the non-Euclidean cases make it plain, first, that the demand *need* not be met by giving an observable thing as the tacit relatum. We may say instead whether or not things move in a non-Euclidean space. But that is not all. Familiar motion, relative to other things in space, does not yield an equally significant idea. It is a weak or recessive sense, bound to yield to the deeper more general requirement that a thing is in motion through space. Thing-thing relative motion might turn out to be a symmetric relation in a suitable theory. But the deeper thing-space motion is asymmetrical on semantic grounds. Space does not move relative to things that it contains (but see section VII, objection 1).

The realist can draw a clear map of the path followed by the post-Newtonian classical physicists. He can relate rest and motion to any of the frames of reference that dynamics and the symmetries of Euclidean geometry together select as indistinguishable. It is at least as clear a map as a relationist can draw. It does nothing to undercut his commitment to a space with indiscernible symmetries. The relativity of motion rests on the commitment. Every frame of reference defines a Euclidean space with the metaphysical properties noted in our list in section I. We may, if we wish, consider these as different hypotheses about what is at rest, what in uniform motion, and the like. Our realism tells us that the difference between these hypotheses cannot matter. There is no clear position that allows us to say that space forms no part of the ontology of post-Newtonian physics. That ontic commitment is there in every choice of frame. It is unclear how relationism avoids begging this question.

I conclude that realism does not entail absolutism.

There is a wild card in this deck, however. It is the metaphysics of continuants. Continuants have spatial parts but not temporal ones. They have histories, but the history of a continuant is a property of it, not a part. Suppose that you and I are continuants. Then I am the person who wrote the paper you are reading: there is no paper-writing part of me nor paper-reading part of you. Neither of us is extended in time. Our histories are our tensed properties, not parts of us. Newton supposed that space was a continuant (though not a substance) and, for this reason, was obliged in general to admit that questions of the identity of places at different times were meaningful and admit that there were different indistinguishable hypotheses to choose (freely) among. A realist need not follow this path. Instead, place classical mechanics in space-time; this will mean that questions about which is the same place at different times cannot arise because space-time is not a continuant entity. Nothing in this bows to detachment theses or spatial relationism: space-time causes much the same consternation among ontologists as space causes and for much the same reasons.

But neither that nor the fact that a realist sees he sometimes cannot care about distinctions between rest and motion means that the questions about rest and uniform motion through space are meaningless. They are idle in the case where space is Euclidean, but only just. Give the space a little uniform curvature and the distinctions engage at once with thing-thing relations. Surely the limiting case of zero curvature inherits the sense of the range of cases of which it is the limit? It is not clear why not. Suppose a space that is changing its curvature from (everywhere constant) positive to (everywhere constant) negative curvature. Do the questions not retain both their point and their meaning as the changing space passes through the limit between the two? I assume that it does.

Suppose we deny the assumption A1. Instead we assert:

A2: We cannot be ontically committed to a space with symmetries.

It would seem to be IP, the indiscernibility or inconsequentiality principle, that might motivate A2. But if that is right, then it would seem to commit us to the view that, though there might well be non-Euclidean space (best, perhaps, space of variable curvature), there could never be a Euclidean space. It is too symmetrical, observationally bland, or anaemic to exist. J. L. Austin once spoke of the idea that existing is like breathing, only quieter. We seem to encounter something like this here, but without the quietism.

Perhaps it is this view which John Earman[6] recommends to a reconstructed ahistorical Leibniz in almost the only place in the philosophical literature where the ideas I have been raising seem to have been mentioned. The relevant two or three sentences come in what appears as a casual afterthought in his splendid paper. Earman points out that Leibniz's arguments fail if the space lacks the relevant Euclidean symmetries. He grants Leibniz anachronistic knowledge of the variety of geometries, and imagines him objecting that Newton used too bland a space to allow him to claim its real existence. Or so I understand Earman. The brief passage is acknowledged and repeated, almost verbatim, by Michael Friedman.[7] This view of things seems to portray Leibniz as insisting on A2 and going on to claim that Euclidean space could not exist, though a non-Euclidean space might.

But I find this suggestion unacceptable. It confuses geometrical differences with ontological differences, or blandness with nonexistence. If Leibniz had indeed raised this objection, surely a decisive reply would have been available to Newton. For, first, neither of them had the least *empirical* grounds for thinking that a non-Euclidean geometry described the structure of space; second, Newton would rightly have claimed that Leibniz was attempting to saddle him with a needlessly complex geometry in order to satisfy an obscure maxim of epistemology which has no place in natural philosophy. Surely the conclusion that there could be non-Euclidean spaces but could not be a Euclidean one is absurd (though not contradictory).

A2 is the child of IP, the inconsequentiality thesis; it leads to absurd consequences, so any version of IP that entails it must be false.

Further Leibniz's argument is fallacious. It is not just a matter of the falsity of his tacit premiss—that every possible space has symmetries of Euclidean space. The light shed by non-Euclidean geometries merely throws into sharp relief an already present appeal to a symmetrical space. Leibniz appeals to that symmetry throughout and it would nonetheless belong to an entity of the sort he wants to deny, but has to presuppose, even if every possible space possessed it.

---

[6] "Who's Afraid of Absolute Space?" *Australasian Journal of Philosophy*, XLVIII (1970): 287–319; esp. p. 303.

[7] *The Foundations of Space-Time Theories* (Princeton: University Press, 1983), p. 221.

Finally, I assume that realism is ill-described as substantivalism. Newton said that he was not a substantivalist and gave a clear reason for saying so. Space has no action on things. Another idea of a substance is simply that of a continuant. Neither of these is entailed by realism but only by specific forms it might take. In General Relativity, mass and energy are features of the space-time metric and affinity. I find it strange that this is described as a relationist view of space and opposed to what is called substantivalism. What have relations to do with it? What else would one want to describe as substantivalist other than the opinion that space-time stores energy and mass, that particles may be generated out of and decay into space-time structure? Certainly not the view that *is* described as substantivalism—that the manifold, despite being undetermined in its structure by the matter tensor, is real. The manifold is too attenuated to be cast as a substance. But it is nevertheless arguably real and the view that it is is surely best so described—it is realism, not substantivalism. Of course, any substantivalist view will make the distinction between container and contained less straightforward than Newton supposed. But that is no feather in the cap of traditional relationism.[8]

VII REPLIES TO SOME OBJECTIONS

1. Let us reconsider nocturnal expansion. When we performed the thought experiment, we did not double strictly everything. We did not double the space. But if we double the space as well as the objects, then there will be no change in thing-thing spatial relations. So is not this Leibnizian argument valid after all?

No. The argument has to yield a particular conclusion, the detachment thesis. It is that thing-thing spatial relations are independent of thing-space spatial relations. But the revised thought experiment does not detach thing-space spatial relations from thing-thing ones. It preserves them. Since it is not a thought experiment in which there is a detachment, it is powerless to defend a detachment thesis.

Perhaps this is a superficial reply, because we have given no meaning to the operation of doubling the size of space. We have yet to give it even a decent syntax—make a well-formed description of it.

---

[8] See John Earman and John Norton, "What Price Substantivalism? The Hole Story," *British Journal of the Philosophy of Science*, XXXVIII (1987). 515–25.

We have not said through what the space is to be doubled. Before we can give even the meager "sense" of well-formedness to doubling our space as well as the things it contains, we must suppose that there is some higher dimensional space that contains it. But a higher dimensional containing space for our space is just an idle fancy. It plays no role in Newtonian mechanics or of any other actual theory of ours. To give to doubling the size of space something we could fairly call a semantic meaning (something stronger than well-formedness) the higher dimensional space has to be given some point in the theory within which discussion is going on. We have yet to see that point. So the idea of doubling space as well as things lacks meaning —so far.

But even if we made the fancy less idle by giving some point to the higher dimensional space, it does not follow that doubling our space through this higher dimensional space would be inconsequential. That depends on whether the wider space is Euclidean. One obvious way to give speculative point to the wider space is to make it non-Euclidean, for then the doubling of our space would change both its shape and that of the objects it contains. The doubling would be consequential. We are still fixated on Euclidean space if we suppose that doubling the space will change no thing-thing spatial relations.

It might seem that the objection has no need of this further space, as my colleague Michael Simpson pointed out to me. The curvature of the space is intrinsic to it, not dependent on some higher dimensional containing space. So it seems that we might double the radius of curvature (i.e., halve the curvature) without referring beyond the space to another which contains it. Then the objection makes sense. We can all agree, however, that metrics equivalent up to a scale factor are not meaningfully different even when the space is finite because of its curvature. This may look like an agreement among relationists and not something to which I can happily turn here. But the point is syntax (addicity) as before: a scale factor can be given sense only when we have something to which to link our space by such a relation as 'Six times as large as . . .'. We can make sense of something being as large as it was only by reference to the quantity of space it filled. As I see it, a relationist ought to insist on this just as firmly as a realist does. The further debate is whether anything could fill the bill as a relatum unless it is somehow privileged (observable, changes thing-thing relations, etc.). We all agree on a point about well-formedness: doubling the space makes no sense as an intrinsic nonrelative operation. The dispute is about whether and how we might give relative sense to it. If the hypothesis of a containing non-Euclidean space through which our space is doubled has consequences for thing-thing relations, the relationist will have difficulty in disallowing that it is meaningful. I am suggesting that whoever raises the specter of somehow doubling the space is, once again, giving me what he hopes is the empty hypothesis of a containing space and defying me to make use of it. But the use is clear.

It is much the same if our Leibnizian fancies fly to the bizarre height of adding to the space of our objects the same velocity vector we added to the objects themselves. Now, when everything moves at a uniform velocity, our space moves, too, along with the objects. The fancy is well-formed only if space-and-objects, holus-bolus, moves through a wider containing space. But, again, if this wider space is non-Euclidean, then both our space and the objects in it will change their shapes, for they will have no parallels in the wider space for the points to move along. The appeal to space as among the things to be changed by any of these thought experiments gets us no closer to the inconsequentiality Leibniz needs.

Contrast this idle fancy with something else. General Relativity is a theory in which tensor equations describe the physical fields. The fields are determined by quantities that might roughly be described as observables. They fix the metric, the affinity, and other things. But they do not determine the lie of these tensor fields on the basic manifold. They do not even fix how the fields lie in local regions beyond which everything is specified. That is the essential thrust of the hole argument. Nevertheless, the manifold is not like the idle fancies just discussed. We cannot define tensor fields except on manifolds: books on tensor analysis begin with manifold theory, as a casual look at chapter headings will show. The same is true of many books on General Relativity. So we cannot begin a tensor theory without first postulating a manifold. General Relativity leaves us with a manifold that has symmetries permitting a variety of different placements of the tensor fields within it. The symmetries show us that it cannot matter which lie on the manifold we choose. The results in terms of observables will not change. But this does not mean that we can take our postulate back and give up realism about manifolds. For, here again, it is in the structure of the manifold—from which the theory must begin—that the symmetries lie. They cannot lie in a nonentity. Nonexistence has no character—in particular, not a symmetrical character. The hole argument fails as an objection to realism. But, again, the realist's perception of the symmetries in the entity he postulates leaves him perfectly free to care nothing about these differences. He plainly sees from the nature of what he postulates that he cannot care about them.

This point is worth making because it may help us to formulate more useful principles about what is meaningful and, rather differently, what to care about in scientific or any other sort of theories. This is a really difficult task. The IP is visibly too strong. Yet we do not want to admit as meaningful every sentence containing some terms semantically anchored in their theoretical role. The preceding example of doubling physical space relative to an idle space meets that weak condition. 'Space' is a semantically anchored term. The idle space is a space in the same sense as physical space, so the sentence about doubling space is syntactically well-formed and thus it is weakly in order semantically. Until we give the idle space a role in theory, we ought to deem the doubling-of-space sentence meaningless, though open to acquiring a meaning. For who can say how theory will advance next? The case of the manifold is different: it certainly has a role in General Relativity. Let those who deny it show us how to get along literally without it instead of merely pointing out its symmetries. But I shall content myself here just with the easier task of concluding that there is much work still to be done in this regard.

2. Here is another objection. Nothing said so far takes us beyond the observation of thing-thing spatial relations. These are all we have under the hypothesis of non-Euclidean space to show us that things would change if doubled over night. Even then, the observations are not enough to establish that the space is non-Euclidean. Poincaré, Reichenbach, Grünbaum, and others have shown with ample clarity that no observations of the sort envisaged can suffice to commit us to a space with any special character whatever.

Reply: Our central question is the argument for the detachment thesis. The issue concerns this conditional: if we change thing-space relations, there are no consequential changes for thing-thing relations. This conditional thesis has been refuted. The objection challenges us to defeat the converse conditional: if thing-thing relations change as envisaged, then they are not consequences of new thing-space relations. But we have no need to refute this. It is not the detachment thesis.

Yet we could refute it. At least, we could argue against it in perfectly familiar ways. The objection throws gratuitous obstacles in the path of reasoning from the evidence to the best explanation. That sort of problem is in no way peculiar to arriving at good geometrical explanation. That is a perfectly general problem for scientific theorizing. We are too familiar (from W. V. Quine and Pierre Duhem) with the way in which tinkering with auxiliary hypotheses and initial conditions can make any body of evidence consistent with ugly, arbitrary, deviant theoretical conclusions. Although the metaphysics of space has proved a happy playground for the invention of obstacle courses, there is no reason to suppose it suffers in a more intimate way from the general malaise. The DP together with the nature of non-Euclidean geometry suggests that the hypothesis of space can do genuine explanatory work (even though it makes no appeal to the action of space). What we have to do here is to argue for the best explanation of the phenomena. Although such arguments are not formally valid, they convince rational people.

Given the evidence we have been speculating about, inference to a space with one or other definite geometry looks like the most elegant and fruitful explanatory hypothesis. We know how to treat space (or space-time) as something that has concrete relations to concrete things and how these changing relations might explain observable changes in the world. General Relativity does this for space-time. But geometrical explanation of concrete changes by means of concrete spatial relations is clear, too.[9] As far as I can see, the main reason why we hesitate to accept geometrical explanations in these cases is that we think the idea of something immaterial and imperceptible—space—is absurd. This does not look like a strong reason. The idea of the material has undergone drastic changes in physics over the last century; the idea of what is observable is anthropocentric and local. It has usually been held that the postulate of a space with a well-defined geometry can do no explanatory work. But it plainly can. Our sense of uneasiness about it begins to look like a vulgar human prejudice.

Is it coherent, meaningful, or intelligible to commit oneself ontologically to an entity that has the sort of stressful combination of properties listed in the introduction? Surely the answer to that modest question has to be 'yes'. It is hard to see what sort of objection can be seriously raised against the intelligibility of entities like space, unless it is that space is neither perceptible nor material. Of course, spatial kinds of things are not perceptible and generally not material. But it is not clear why these should be regarded as defeating the commitment as long as space has well-defined characteristics, is a particular, has concrete relations to concrete things, and can explain concrete changes in things (such as change of shape after doubling). If space can deliver these and other sorts of theoretical goods for us, it does look like prejudice to insist on materiality and perceptibility. And clearly Leibniz shows us how well we can understand the characteristics of a space by appealing so ingeniously to a Euclidean character for it.

---

[9] See my "What Can Geometry Explain?" for illustration and discussion of this.

Could it ever be useful to include space in the ontology of some theory, or would we always be better off trying to avoid the postulate? The answer to this question is much less clear, though my own view is that space does some work in easily imagined theories, and does it elegantly enough to earn its keep. Relationist alternatives, by comparison, rob theoretical language of much of its assertive force and burden it with awkward conventions and indeterminacies. They impede a clear view of what the theory does, clutter it with nonfactual elements while offering no gain beyond the economy that they do not postulate space. Even then, I have been arguing, it is unclear how they really do avoid the postulate.

Leibniz's argument has loomed large because relationism needs to argue for the detachment thesis. The challenge to Clarke was no mere historical episode in which we now take only an antiquarian interest. It remains the most attractive relationist gambit. Relationism has no point without the detachment thesis. The argument of the challenge is fallacious, begging the question. On the other hand, if the discernibility principle is true, we have a strong argument from it to the meaningfulness of thing-space relations generally and so of space itself. It is easy to see how we might arrive at a posit of non-Euclidean space as our best explanation of what we observe.

GRAHAM NERLICH

University of Adelaide

AGENCY AND MORALITY*

AN "agent-centered restriction" (ACR), according to Samuel Scheffler,[1] is "a restriction which it is at least sometimes impermissible to violate in circumstances where a violation would serve to minimize overall violations of the very same restriction, and would have no other morally relevant consequences" (*ibid.*, p. 409). It would be wrong, for example, intentionally to kill one innocent person simply to minimize the number of innocent people intentionally killed. Similarly, breaking a promise, violating a right, or committing an injustice cannot be justified simply to minimize the number of identical broken promises, rights violations, or instances of injustice.

Because of this character, such restrictions harbor, for Scheffler, a "persistent air of paradox."[2] If a particular action is reprehensible, how can it (he suggests) be wrong to do it if, without other moral consequence, this will minimize the number of such actions. He notes that, although ACRs are central to much of our common morality, they apparently deny what seems to be the quite rational goal of minimizing what is morally objectionable.

Scheffler thinks this alleged conflict between our common (deontological) morality and a claim of reason creates a special burden for deontology; for, unlike ordinary consequentialism (e.g., utilitarianism), deontology rejects the principle that the moral good is to be maximized. We are not enjoined by its prescriptions to produce the most welfare, or right's protection, or virtue, or anything else (*ibid.*). And yet, how can it be wrong to torture one child (my example), if by so doing one simply prevents the identical torture of two children.[3]

---

* I would like to thank members of the Eastern Pennsylvania Philosophy Association and of the Philosophy Club of the University of St. Andrews, and especially my late colleague Seymour Schwimmer for help on earlier drafts of this paper.

[1] "Agent-Centered Restrictions, Rationality and the Virtues," *Mind*, XCIV (July 1984) Also see *The Rejection of Consequentialism* (New York: Oxford, 1982), pp. 80, 112 The nonmaximizing character of such restrictions is central It is this property I am generally referring to in the paper when designating a moral point of view as "deontological," "nonconsequentialist," or "our common (or ordinary) morality." This is not meant to slight those personal commitments having to do with family, friends, etc They certainly are a part of our "common morality " My aim, though, is to focus, as Scheffler does, on that powerful belief that it can be wrong to cause harm to prevent even greater harm

[2] *The Rejection of Consequentialism*, p. 112.

[3] Scheffler apparently assumes about welfare or anything else conceived from a moral point of view that, if a particular action or consequence is bad, then, other things being equal, more instances are worse. Instance aggregation, however, though essentially unquestioned by Scheffler, may not be crucial to his challenge to

0022-362X/91/8804/190–212

Whatever makes torture objectionable, the torture of two children is worse than the torture of one. Moreover, if I torture a child, it is not, for that reason, worse than someone else torturing a child. Why should it matter (so the challenge goes) that one instance is caused by me if, without other moral consequence, that minimized instances of torture?

Scheffler, then, presents a striking internal challenge to deontology; 'internal' because of the suggestion that any moral perspective, including deontology, that identifies what is objectionable has reason to be an overridingly maximizing perspective. It would be as if to have a coherent conception of the decalogue one required an implicit eleventh commandment (or at least permission) that one could kill, steal, etc., in order at least to minimize the instances of killing, theft, etc. My aim in this paper is to answer the challenge by arguing for three claims: (I) there are three quite different conceptions of the importance of agency to moral obligation that should be, but often are not, kept distinct; (II) the particular sense of agency that Scheffler and others appear to believe is embodied in ACRs is irrelevant in accounting for their intuitive power; and (III) a patient- rather than agency-based defense of such restrictions is plausible and in accord with what, for Scheffler, is an important principle of practical reason. At the end of the paper, I make some further comments about agency in the light of this defense.

I AGENCY AND MORAL OBLIGATION

Assuming the label 'agent-centered' (or 'agent-relative') is not simply a name for a nonmaximizing morality, we might wonder in what sense (if any) such a morality is connected to agency.[4] After all,

---

deontology. There are certainly cases where common morality forbids doing something to one person, although if you refrain, something worse is done to someone else. One cannot cut off the finger of one child to prevent the torture and death of another. In this paper, I accept for the sake of argument (with private reservations) that, if a particular action or state of affairs is morally objectionable, then more instances of the same are more objectionable. On this question of aggregation, see John Taurek, "Should the Numbers Count," *Philosophy and Public Affairs*, VI, 6 (Summer 1977): 293–316; Brook and Schwimmer, "On Adding the Good," *Social Theory and Practice*, VII, 3 (Fall 1981)· 325–37; and Frances Myrna Kamm, "Equal Treatment and Equal Chances," *Philosophy and Public Affairs*, XIV, 2 (Spring 1985). 177–95.

[4] In some cases, Scheffler appears simply to stipulate that "agent-centered" is meant just to designate a nonmaximizing morality. "The agent relativity of these constraints and permissions consists in the fact that their application to individual agents is not based on or supported by any appeal to what would be best from an impersonal standpoint"—Scheffler, ed , *Consequentialism and Its Critics* (New York: Oxford, 1988), p. 5. But even if only a name, we would still want to know why 'agent-relative' was the appropriate one. And in a recent response to Jonathan Bennett on this question, Scheffler gives a more substantive (though I think mislead-

restrictions against harming some to aid (even more) others appear victim-centered, particularly if they are based (as I argue for later) on ascribing certain rights to people. In deciding whether or not to honor a right, I may have no concern that *I* rather than someone else might violate it. In his Tanner Lectures, Thomas Nagel[5] appears to conflate these considerations. He suggests rightly that deontological constraints begin with "claims of other persons not to be mistreated." Yet, against Nagel, it would seem precisely these claims which refer to agent-neutral values, since they concern states of affairs that give, in his words, "a reason for anyone to want or not to want" (e.g., not to have one's rights violated). Nagel says, however, that such claims give an agent the "agent-relative" reason "*not to maltreat others himself.*" Presumably, this reflexive use of the pronoun 'himself' is not simply redundant; what it suggests is that, for the agent, it is particularly important in honoring a claim that *he* not maltreat others rather than simply see to it that others are not maltreated (or, more broadly, not harmed).

This view about the significance of agency, which I call 'A1,' is given brief expression by Bernard Williams[6] when he says that "each of us is specially responsible for what he does rather than for what other people do" (*ibid.*, p. 117). Recently, Stephen Darwall[7] makes the comparable claim when discussing what he calls "agent-relative obligations":

> It is unquestionable that moral common sense suppose there to be *prima facie* obligations that are agent relative. The obligation not to

---

ing) answer to why deontological constraints have some significant relation to agency. I comment on this answer in the text. See Bennett, "Two Departures From Consequentialism" and Scheffler, "Deontology and the Agent, A Reply to Bennett," in *Ethics*, C, 1 (October 1989): 54–67 and 67–76, respectively. Jeremy Waldron, commenting on Robert Nozick's notion of "side constraints," suggests these constraints are "agent-relative" because each agent is taken to be concerned only with her own observance of the constraints"—"Rights in Conflict," *Ethics*, XCIX, 3 (April 1989), p. 504. Literally, of course, this is false (as Waldron would undoubtedly admit) since we can certainly have obligations to prevent others from violating side constraints. Waldron's implicit suggestion, however, and the one I take up in the essay, is that calling a restriction "agent relative" means at bottom that one has a greater responsibility not to violate a constraint than to prevent someone else from violating the identical constraint (or preventing a comparable natural harm). I argue in the paper that this contrast is misleading because, although true, it is a derivative rather than a central point about deontology

[5] "The Limits of Objectivity," reprinted in James P. Sterba, ed., *The Ethics of War and Nuclear Deterrence* (Belmont, CA: Wadsworth, 1985), pp. 15–23.

[6] "Utilitarianism and Integrity," in J. J. C. Smart and Williams, *Utilitarianism, for and against* (New York: Cambridge, 1973), pp. 76–117.

[7] "How Nowhere Can You Get and Do Ethics," *Ethics*, XCVIII, 1 (October 1987): 133–57.

> betray a friend, for example, cannot be reduced to the neutral value of betrayals. If it could, it would be morally indifferent, other things being equal, whether one betrays a friend or does something that results, however indirectly, in a comparable betrayal in which *one is not directly involved at all* (my emphasis) (*ibid.*, p. 150).

Unlike Williams, Darwall is discussing the special duties of friendship, but his point is the same: although my betrayal of a friend is from an impersonal standpoint no worse than someone else's betrayal of a friend, I am more responsible not to cause that harm (betrayal) than to prevent others from causing an identical harm ("to prevent harms I am not involved in"). And, indeed, this claim that I am more responsible for my own actions than for the identical actions of others might, if correct, support the view that it is wrong to torture one child to prevent someone else from torturing two. A1 conceives moral obligation, then, in a personally indexed way, implying that the requirement that *one* not violate moral restrictions is more stringent than the demand simply to prevent violations.[8] A1 is, of course, too narrowly conceived if the contrast is only between my harming and my failure to prevent others from harming. Deontological restrictions, broadly conceived, forbid my harming to nullify threats initiated either by others or by nature. And, of course, natural threats are not violations.[9] Rightly conceived, then, A1 is the view that the duty not to cause harm is more stringent than the duty to prevent comparable harm.

The second sense of agency used to make a moral distinction, A2, is fundamental in the usual expression of deontological restrictions and is formulated (roughly) by the "doctrine of double effect": it is impermissible to cause a particular harm *in order* that (even more) good might occur, although it may be permissible in realizing that good knowingly but unintentionally to cause an identical harm. The doctrine has immense problems with respect both to its formulation and application, but it or some analogous distinction is historically central in expressing the nonmaximizing core of common morality.[10]

---

[8] It is striking how this conception of the importance of agency to deontology has gained currency in contemporary discussion. In a recent introductory text, we find the following: "An agent-centered pluralist morality could agree that deceit, coercion and injustice are bad. But it differs from consequentialism in giving each agent different aims; each of us should have the aim that *he* should not coerce or deceive other people, even if . . . [this] would make it the case that there was less coercion overall than there otherwise would have been. I shall follow tradition in describing agent-centered pluralist moral systems as *deontological* theories" (author's emphasis)—David McNaughton, *Moral Vision* (New York: Blackwell, 1988), p. 166. I mention later in a footnote an uncontentious sense in which deontology might be said to possibly give each agent different aims, while in another and obvious sense it gives each agent identical aims (see (fn. 21).

[9] Bennett (*op. cit.*) properly makes this point. Williams is, however, interested in the narrower contrast between responsibilities for my own actions and responsibilities for the free actions of others. This is the contrast raised in blackmail cases when I am asked to do evil to prevent someone else from freely doing more evil. I think these cases raise significant, distinct, and difficult questions, but Bennett rightly considers them too narrow to capture intuitively the restriction against harming some to prevent harm threatened by others or by nature.

The third sense of agency, A3, is a temporally indexed version of the doing-allowing distinction. It claims that *at a given moment* one is more responsible not to cause harm than to allow harm to occur. A1 and A3 are, unfortunately (as I argue below) not always distinguished in the usual puzzle cases about harming some to aid others.[11]

Clearly, A1 and A2 are quite different. Double effect draws a distinction within the compass of one's own agency; it permits balancing the good and bad effects of one's acts only if the foreseen but bad effects are unintended. There is perhaps something paradoxical here, but it is the paradox of distinguishing what is forbidden and permitted *for the same agent* by whether what is done is intended or only foreseen. Why should it be more permissible to discount the ends of another in foreseeing her harm as a consequence of one's action, than to discount the ends of another in merely using her for one's purpose? However we puzzle this out, A1 apparently ignores intention but, perhaps as paradoxically, suggests that, although a particular action is equally wrong whether done by *oneself* or *others*, one is more obligated not to do it than to prevent others from doing it.

---

[10] The Kantian distinction between using or not using someone as a mere means is a related distinction. This appears to be distinct from double effect, but see Onora O'Neill, "Kantian Approaches to some Famine Problems," in *Matters of Life and Death*, T. Regan, ed (New York: Random House, 1980), pp. 285–94. O'Neill argues that to use persons as mere means is equivalent to intending to treat them in a way incompatible with their interests, therefore intending to harm them. For an interesting discussion of a set of distinctions with a family resemblance to double effect, see Christopher Boorse and Roy A. Sorensen, "Ducking Harm," this JOURNAL, LXXXV, 3 (March 1988): 115–35. See also the recent attempt by Warren Quinn to capture a more generic sense of mistreatment that is meant to include both strictly intending harm to a person and merely using a person which foreseeably harms her—"Actions, Intentions and Consequences," *Philosophy and Public Affairs*, XVIII, 4 (Fall 1989): 334–52 There is a great deal of literature, both religious and secular, on the doctrine of double effect, but it is not the main topic of this paper. Some other recent interesting discussions are in Charles Fried, *Right and Wrong* (Cambridge Harvard, 1978); Nagel, *The View From Nowhere* (New York. Oxford, 1986), ch. IX; and Nancy Davis, "The Doctrine of Double Effect: Problems of Interpretation," *Pacific Philosophical Quarterly*, LXV (1984)· 107–23. Most of this literature attempts some justification for a strong restriction on intentional harm. Very little of it offers a needed defense for the central claim of the doctrine that, in principle, merely foreseeable harm is permissible. Fried's work is an exception

[11] Phillipa Foot expresses something like A3 when she says: "Nevertheless it is from the moral point of view a different thing to start up a new train of events in someone else's life and to refuse to intervene to stop one", "Morality, Action and Outcome," in *Morality and Objectivity*, Ted Honderich, ed. (Boston Routledge & Kegan Paul, 1985), p. 24. Since there is no temporal index in Foot's formulation, it is not clear she distinguishes the doing-allowing contrast from the responsibility for self versus responsibility for others contrast. For an interesting discussion of ambiguities in an earlier paper of Foot's, see Nancy Davis, "The Priority of Avoiding Harm," in Bonnie Steinbock, ed., *Killing and Letting Die* (Englewood Cliffs, NJ· Prentice Hall, 1980), pp. 173–215. Foot's earlier piece is "Abortion and the Doctrine of Double Effect," repr. in J. Rachels, ed , *Moral Problems* (New York. Harper & Row, 1975), pp. 59–71.

A1 and A2 can be ambiguously lumped together. Take, for example, the following illustration of Nagel's: "And it seems you shouldn't twist the arm of a small child *to get* its grandmother to do something, even if the thing is quite important—important enough so that it would not be reasonable to forgo a comparable benefit in order to prevent *someone else* from twisting a child's arm" (my emphasis).[12] Both senses of agency are suggested: intention (twisting the arm to get a result) and the alleged special responsibility for one's own actions as opposed to the actions of others (not twisting the arm yourself as opposed to preventing someone else from twisting the arm). Yet Nagel ultimately opts for the distinction between intention and mere foresight as embodying what he calls "agent-relative values," values that express, as he says, the "particular relation of the agent to the outcome" (*ibid.*, p. 17). He contends that the simply foreseen consequences of one's actions [are] "not in the *relevant sense* what one *does*, and [do] not come under a deontological constraint" (*ibid.*, p. 19; my emphasis).

This claim that the foreseen consequences of my actions are not "in the relevant [moral] sense" what I do assumes without argument (here at least), however, that such consequences should not be subject to constraints as strict as those on consequences intended. That point would not automatically follow even if it could be made out that instrumentally harming was intrinsically worse than consequentially harming. Double effect is at bottom a doctrine about what is permissible. Therefore, to make his claim, it is not sufficient for Nagel to remark on some special reprehensibleness in "aiming" at evil (e.g., the child's pain) as an instrument for good.[13] This point, though suggestive, tells us little why, if unintended, we can permissibly cause an identical harm.

---

[12] "Agent Relative Morality," in *The Ethics of War and Nuclear Deterrence*, p. 19.

[13] *The View From Nowhere*, pp. 176–80. Nagel's point is that, in intending to cause the child pain, one "tracks" the pain, twisting the arm harder to increase it if that is needed. But we can imagine something quite similar about unintended but foreseeable harm. For example, suppose that, if we warn civilians who might be killed as a collateral but unintended effect of bombing a military installation, the installation would be moved. So we cannot warn the civilians. We might, in fact, do everything to ensure they stay. We in effect track the civilians and ensure that they die, though we do not intend their deaths. In principle (subject to proportionality constraints), double effect permits this.

In any event, by labeling deontological restrictions 'agent-centered', Scheffler is not merely referring to double effect's distinction between intention and foresight. His challenge is to the rationality of forbidding one violation of a restriction though that serves, without other consequence, to minimize identical violations. Double effect, however, marks the primary moral difference between types of things one does rather than between consequences caused and consequences not prevented. Scheffler's examples and discussion strongly suggest that the sense of agency he thinks embodied in ACRs is A1.[14] Commenting on the "ordinary view" of the contrast between negative and positive duties, he says: "But if *you* would have to murder one innocent person to prevent the murder of five by someone else then you must not do so."[15] And in the following passage, he gives "a schematic example" of the alleged paradoxical character of ACRs. "Suppose that if agent A1 fails to violate a restriction R by harming some undeserving person P1, *then five other agents A2 . . . A6,* will each violate restriction R by identically harming five other persons, P2 . . . P6, who are just as undeserving as P1, and whom it would be just as undesirable from an impersonal standpoint to have harmed" (*ibid.*, p. 84; my emphasis). This suggests that Scheffler, like Williams, Darwall, and Nagel, thinks ACRs involve a personally indexed view of the importance of agency, i.e., that the agent conceives her special obligation that *she* not do wrong (as opposed to just seeing that wrong not be done), where the contrast emphasized is between obligations for one's own actions and obligations for the comparable actions of others.[16] Scheffler makes the same point in a recent exchange with Jonathan Bennett. Discussing the "doing-allowing" distinction, he correctly notes that that distinction (or Bennett's "making-allowing" distinction) cannot be fundamental to deontological restrictions. If it were, consequentialism gains support, for then it seems plausible to permit doing *X* (or making *X* happen) if only to reduce the number of doings (or makings) of *X*. Scheffler puts it this way: "the significance of that distinction [doing-allowing] is agent-relative in character; what matters are the agent's doings, not the total number of doings overall."[17]

---

[14] Scheffler, of course, is not defending A1. As a matter of fact, he is quite skeptical about its worth. I note, as well, that double effect's strong proscription against intentionally causing harm would be—using Scheffler's terminology—a species of agent-centered restriction, since one is forbidden to harm intentionally even to prevent more intentional harm.

[15] *The Rejection of Consequentialism*, p. 24. The italicized 'you' is interesting. There is a difference between: (1) *you* should not kill; and (2) you should not *kill*. The first appears to make moral restrictions relative to the agent. The second more plausibly expresses the general prohibition against killing which distributively applies to any agent.

[16] This is the way Christine Korsgaard interprets Scheffler in "Personal Identity and the Unity of Agency: A Kantian Response to Parfit," *Philosophy and Public Affairs*, XVIII, 2 (Spring 1989), p. 102. An agent-centered morality, she writes, requires that "you must be in a special way concerned with what *you* do."

Presumably, Scheffler means that what matters particularly to the agent is the agent's doings as opposed to the doings of others.[18] And it is this conception of differential responsibility which I am labeling 'A1'. Yet, significant as A1 appears to be in accounting for our ordinary (deontological) moral intuitions, I argue below that it is derivative rather than central to the force of those intuitions and therefore cannot be central to any justification for deontological constraints.

II IS A1 ESSENTIAL TO DEONTOLOGY?

Consider the following case. You are at the zoo with two children who are making a scene. Becoming angry, you toss them into the lion's den. Horrified, you come to your senses and notice that they can only be saved if you toss a third child (who just toddled along) into the back of the den. The beast would be distracted and you could quickly leap in and save the first two. What should you do? Whatever you do, you have initiated a chain of events that results in a child's death. If you do nothing, you have intentionally killed the first two children. If you sacrifice the third child, you kill it in using it as a mere means to save the others. Or consider this variation of Darwall's example about betrayal. Captured by the enemy and under some pressure, I reveal the hiding place of two comrades. They will be killed if discovered. I could prevent this by revealing the whereabouts of a third comrade, perhaps of more importance to the enemy. Whether or not I reveal her whereabouts, I have intentionally betrayed someone.[19] And equally important, if I fail to betray the third comrade *I still remain directly involved* in the betrayal of the first two.

---

[17] "Deontology and the Agent," *Ethics*, c, 1 (October 1989), p. 73.

[18] There is, of course, a minimum and uncontentious sense in which deontology may give each agent different aims. The state of affairs described as 'Smith not intentionally killing an innocent person' can refer to an aim of Smiths but not necessarily to an aim of Jones. The personal index is read into the content of the aim. Ordinarily, however, the personal index is read as an operator—each person has the identical aim not to kill an innocent person. Having the index in the content of the aim suggests, misleadingly, that a belief about differential responsibility—that one has more responsibility for one's own actions than the comparable actions of others—is central to deontological restrictions. But that it is central (as opposed to being derivative) certainly does not follow from reading the personal index into the aim. That reading leaves it open what Jones's responsibility is with respect to the aim described as 'Smith not kill an innocent person'.

[19] Of course, in this example more than one agent is involved. The important identity of structure in the examples, however, is that the person who initiated the

Yet it seems clearly wrong to kill the third child or betray the third friend; as wrong, in fact, as in any of the standard cases where you might harm one person to aid others. In terms of what is important, these are, in fact, standard cases. It is perhaps just an artifact of the usual examples that agency (in the sense of A1) appears to be important. Perhaps we are more likely to face harming some to nullify threats initiated by others (or by nature) rather than by ourselves. But that difference in the origin of the threat is irrelevant to the intuitive force of the prohibition against harming some to aid others.

Of course, A3 might justify not killing the third child or betraying the third friend. One can claim the agent has a greater responsibility *at a given moment* not to violate a constraint than *at that moment* to prevent a violation. The temporal index here, however, is crucial. Without it, the doing-allowing distinction cannot provide a rationale for not killing the third child. Whether or not I kill that child, I have done something, killed children, rather than merely allowed them to die. Admittedly, Scheffler's definition of an ACR on occasion appears compatible with either A1 or A3.[20] Yet the latter, I think, fails to connect agency in any significant way with agent centeredness; that is, it does not connect moral restrictions with one's agency interpreted as a special concern with one's own doings overall. After all, what counts as what I do extends over more of my lifetime than to what I do at a particular moment. It is true that the temporally indexed version of the doing-allowing distinction would, in accord with common morality, forbid my murdering one child though that *allows* it to be the case that I have murdered two others. But A3 is far from being "agent-centered" in the sense of giving me a special obligation that *I* not violate deontological constraints. If a restriction were truly agent-centered in this sense, each agent would have the plausible goal of minimizing her own violations as opposed to total violations.[21] In the case of the children, however, that is precisely what deontology forbids. It is wrong to murder the third child to nullify my previous attempt to murder two. Single-agent examples (where the person who can nullify a threat, originated the threat)

---

threat is the same person who might nullify it by harming someone else Under A1, in my version of Darwall's example, I would have the same responsibility for the fate of any comrade since after the first betrayal I am responsible for a death whatever I do

[20] For example, the passage quoted at the beginning of this paper. And, as I mentioned (fn. 4), he suggests in a recent piece that 'agent-relative' simply designates a nonmaximizing morality. But this raises again the question why, in a substantive sense, nonconsequentialism is an agent-relative point of view.

[21] I discuss later in the paper the view that intention rather than consequence determines wrongness. In not doing that here I think I am being fair to the spirit of Scheffler's notion of constraint violations

allows, in fact, a criticism of Scheffler similar to his criticism of Bennett. Scheffler's point against Bennett was that, if the doing-allowing distinction were fundamental, one reasonably should be permitted to do *X* to minimize the total number of doings of *X*. But in the same way, if an agent were overridingly concerned with her own doings, a limited consequentialism equally gains a foothold. The agent plausibly should be permitted to do *X* to minimize her own doings of *X*. But, again, deontology forbids this.

We might combine both the temporal and the personal indexing and say the agent, at a given moment, has a special obligation that she (at that moment) not cause a violation. But the two forms of indexing work at cross purposes. My personal responsibility again is substantially coextensive with my identity over time as an agent. At any given moment *t*, I am as responsible for the consequences of actions I initiated prior to *t* as for those I begin at *t*.[22] The temporally indexed version of the doing-allowing distinction, however, discounts the general significance of my agency since, at a given moment, it makes me more responsible not to violate a moral restriction then to see to it (at that moment) that *I* do not end up violating more restrictions. I do think (and defend later) that A3 remains significant for deontology, but it is of derivative rather than primary importance.

In any event, the personally indexed sense of agency, A1, which appears to be what Scheffler means by "agent centeredness," cannot account for the conviction over the full spectrum of examples that it is wrong to cause serious harm to some to nullify threats of more harm to others; for the compelling power of this conviction remains when one considers cases where one is the only agent. In these cases, no contrast is made between my obligations not to harm and my obligations to prevent others (or nature) from harming. Yet it is, for ordinary morality, as wrong for me to torture one person to nullify either my own or another's previous threat to torture two.[23]

What remains distinctive about deontology, then, is not some personally indexed view of moral responsibility, but simply its nonmaximizing character. And, indeed, this character is perhaps even more paradoxical (as Scheffler[24] notes) if A1 is rejected. For if it is morally irrelevant who (or what) brings about a consequence, it might seem that only the consequence matters. Of course, the irrelevancy of A1 does not rule out an agency-based (if not agent-centered) defense of common morality using either a temporally indexed version of double effect or of the doing-allowing distinction.[25] Another possibility, however, is a patient-based defense of deontology. This is the direction I take in the last section of the paper.

---

[22] This should be amended somewhat. If an agent radically changes personality, we might say she is no longer morally responsible (as opposed to physically responsible) for prior actions. But in any case my moral responsibility, though perhaps not coextensive with my physical agency, includes more than the consequences of actions I initiate in the present.

[23] Although the stuff perhaps of science fiction, we could imagine that Smith has started a device that, if not stopped, will torture two people. Only the scream of a third person, induced by a certain amount of pain, has the required pitch to destroy the device. I owe the idea for this type of example to a discussion of double effect with Nozick.

I would note first that, in responding to some criticism by Phillipa Foot, Scheffler denies that his challenge requires some conception of an "overall best or worst state of affairs," as, for example, in utilitarianism. Rather, he contends that consequentialism embodies a deep and quite general principle of rationality which he calls "maximizing rationality" (MR):

> The core of this conception of rationality is the idea that if one accepts the desirability of a certain goal being achieved, and if one had the choice between two options, one of which is certain to accomplish the goal better than the other, then it is, *ceteris paribus*, rational to choose the former over the latter.[26]

Agent-centered restrictions appear to violate this principle because they claim certain states of affairs are bad (e.g., a child being tortured), but (*ceteris paribus*) forbid acting so as to minimize the number of such states. "How," Scheffler asks, "can it be rational to forbid the performance of a morally objectionable action that will have the effect of minimizing the total number of comparably objectionable actions that are performed and have no other morally relevant consequences?" (*ibid.*, p. 413).

Scheffler allows that deontological restrictions could be substantially compatible with MR but only if the objectionableness of an action's consequence does not exhaustively account for the restriction (*ibid.*, p. 418).[27] It is interesting that the additional element he thinks might be required for compatibility appears to be something like A1. He suggests, for example, that, if for each agent, the fact that *she* was violating a restriction was an element in the objectionableness of the violation, then each agent presumably might have the coherent agent-relative goal that she not violate a restriction (as opposed to the goal of minimizing the total number of violations). This is not a view which he thinks credible, since he thinks it conflicts with the impersonal character of moral aims, but a view, he believes, which would at least make deontological restrictions consistent with MR: "views that include agent-centered restrictions need not come into conflict with maximizing rationality when they tell us to further the agent relative goal of not violating the restrictions *ourselves* at the expense of the non-relative goal of minimizing violations of the restrictions" (*ibid.*, p. 417; my emphasis).[28]

---

[24] "Agent-Centered Restrictions, Rationality and the Virtues," p 414.

[25] In order to work in single-agent examples, double effect would also have to be temporally indexed, i.e., what at a given moment I intend or simply foresee. In the children and the lion example, I violate the nonindexed version of double effect whether or not I kill the third child.

[26] "Agent-Centered Restrictions, Rationality and the Virtues," p. 414. Foot's discussion is in "Utilitarianism and the Virtues" and "Morality, Action and Outcome," in *Morality and Objectivity*.

[27] I say 'substantially' to put aside the truism that, if an agent's goal at a moment is not to violate moral restrictions, then MR is automatically satisfied by her acting in accord with deontology's constraints. But this response masks the significance of Scheffler's challenge that, once the agent's goal is given a more substantial description, e.g., 'don't use anyone merely as a means', our attention is directed to a particular consequence (being used in a certain way). There then seems some intuitive pressure to minimize such consequences

Against Scheffler, I have argued that A1 (though perhaps in accord with MR) is not compatible with restrictions against harming some to aid others. An agent taking seriously A1's view of moral responsibility could reasonably choose to minimize her own total violations (as opposed to all violations), and thus to do what ordinary morality forbids, kill one person to nullify her prior attempt to kill two.

Scheffler, then, finds the constraints of ordinary morality paradoxical, because he apparently thinks A1 is required for them to be consistent with MR, but believes A1 to be unjustified (*ibid.*, p. 415).[29] I do agree that A1 appears to conflict with the impersonal character of moral requirements. It is no element (by itself) in the wrongness of an act that I do it rather than someone else. But, of perhaps more interest, A1 is not even compatible with deontological constraints since it appears to permit violating a constraint to reduce one's total violations of the same constraint. In fact, it is this failure of A1 to support nonconsequentialism which suggests the latter needs a patient- rather than agency-based defense.

III. A PATIENT-BASED DEFENSE OF DEONTOLOGY

To answer Scheffler's challenge, we must locate some property *P* of a patient for which the following propositions are true: (1) the existence of *P* explains why a state of affairs characterized by *P* is morally objectionable; (2) if a subject *S* ends up bringing about two instances of *P*, she has done (in terms of the consequences of her act)[30] something worse than another subject *S** who brings about one instance of *P* (*S* can be identical to *S**); and (3) the impermissibility of causing one instance of *P* simply to prevent more instances of *P* is in accord with practical reason. Take the case, again, where I consider killing one child by using its body to distract a lion from two I previously endangered. Regardless of who endangered the first two, it seems quite wrong to kill the third child. What might account for the force of this intuition? The case can be tightened up by making it "one on one." Assume I tossed one child into the den, but could save him by distracting the lion with a second (who then would be killed). Whether I sacrifice the second child or not, I have unjustly killed a child. Having no concern now about the numbers, one might seek some property possessed by the second child to defend a principled refusal to kill it.

---

[28] Scheffler argues here, persuasively I think, that this personally indexed sense of agency is not compatible with taking "the moral point of view." This appears to be the basis for his view that there is some "*prima facie* irrationality" in deontology; *Rejection of Consequentialism*, p. 83.

[29] The point he makes here is that the moral (as opposed to the egoist) point of view must take a stand about the objectionableness of states of affairs (say a child being tortured) independent of agency

The obviously distinct property of the second child is that it has not at that moment been significantly threatened.[31] I mean by this at least (and somewhat roughly) that no causal chain has been initiated which, if uninterrupted, will imminently lead to its harm.[32] The second child might well have a disease that will inevitably kill him in twenty years. But this does not alter his status as significantly unthreatened when I consider using his body to distract the lion. Another way to put this is to say that, even with the disease, the second child is significantly benefited by my not injuring or killing him to save the first.[33] As an imaginative device, we might think of a moral physics where the life lines of people pursuing their interests but not threatening others are analogous to force-free or inertial motions in mechanics. As in a natural physics no explanation is required for inertial motion, so here no justification is required for individuals to pursue their nonthreatening interests. And just as in mechanics deviation from inertial motion requires an explanation, distorting the path of an individual by threat requires justification. There is a moral shield, possessed by each person, a shield against threats, i.e., actions that change the world so as to bring an individual into harm's way. The fundamental notion is that not being significantly threatened at a given moment is a basic and morally important property.[34] The child I might kill to save the first possesses at that moment the property of not being imminently threatened with death, a property lacked (then) by the first child.

---

[30] It is important to emphasize that (2) only refers to consequences of an action. And by 'consequence' here I mean anything that could be the referent for a consequence description, e.g., 'harming a person', or 'violating a person's right'. There is, of course, a dimension of "wrongdoing," that allows grading actions in terms of which is worse than another in terms of *mens rea* rather than consequences. I discuss this later in the paper as a possible response to Scheffler. This note was prompted by some comments of an editor of this JOURNAL.

[31] Except for the Pickwickian sense in which the second child is threatened because I might decide consequentialism is correct.

[32] A difficulty here is how to consider sincere threats from, say, terrorists. We might not want to consider such threats as themselves the initiation of causal chains that, if uninterrupted, will lead to harm. In the case I am considering, however, the threat is already embodied in an attempt at harm.

[33] What of cases where, if you did not kill a person to save others, that person and the others would be immediately killed by someone else? That person is not benefited by your refusal to kill him. Williams pursues this kind of case, suggesting that the moral importance we place on our agency accounts for the belief that it may still be wrong to kill the one. Williams, however, does not make much of the aspect of overdetermination, i.e., that the one you might kill would die in any event. His point is to contrast one's responsibility for one's own actions with responsibility for the free actions of others. Thus, it is not clear to me how Williams would think of examples where either (a) the one would not die if you did not kill him, or (b) if you did not kill one, all would be killed by natural causes (or a process already initiated by another); see "Utilitarianism and Integrity."

It is important in this regard to remember that Scheffler's challenge to deontology contains a *ceteris paribus* clause: he suggests that, if a state of affairs, *S*, is objectionable and other morally relevant considerations are equal, it seems reasonable to minimize the instances of *S*. Certainly, the clause is satisfied if my choice is simply to prevent one murder or two. Then (allowing adding the value of lives) I should prevent two. On the other hand, if I have to murder one child to prevent my having murdered another, then the *ceteris paribus* clause is violated. But not—and this I think quite important —because of the moral significance of my agency. What is significant is not that I (as opposed to someone else, or some natural cause) would cause the death but rather the protection due the child who at that moment is not threatened.

The morally significant feature of not being threatened at a moment demands some special protection, a protection embodied in the right not to be harmed. Expressing the right as one simply against threats (as opposed to the consequences of threats) appears too weak and is in fact unnecessary, since, except for special (though interesting) cases, harming a person entails having altered the world in a way that significantly threatens her with harm.[35] Thus, the right not to be harmed makes concrete the special protection against threats of harm.

---

[34] Obviously, there is some idealization here. A person might be unthreatened at a moment but deserve harm for some past action; or the person might simultaneously be unthreatened and wrongly threatening others; or he might be innocently threatening others. I exclude such cases as we might ignore various disturbances in an idealized natural physics. This is not to deny their importance.

[35] The exceptions I have in mind are cases of overdetermination, e.g., drowning a person who has fallen out of a boat, or stealing money from a person who is fleeing a thief. These seem to be examples where harming a person does not entail originating the threat (at least in a generic sense) of harm. But in the standard puzzle cases where harming some is required to aid (more) others, the act of harming originates in some significant sense the threat of harm.

But however we think of the content of rights, what is important is that a protection expressed as a right requires a principle of "non-trade-off," i.e., one cannot violate a right simply to minimize the number of identical right's violations. Although (some if not all) rights can be overridden, minimizing the number of identical violations is not, in itself, a sufficient reason to violate a right.[36] Although the "non-trade-off" property is an accepted feature of rights, it is not just being stipulated here. In the case of the children, the non-trade-off feature expresses the special protection due the then-unthreatened child. If no special protection were due that child, it would be immaterial in the "one on one" case, other moral considerations being equal, whether I killed the second child to save the first, or did not kill the second child and thereby killed the first (I could flip a coin). Either way, I minimize violations of the right not to be killed. Contrapositively, if some special protection is due the unthreatened child, then it is wrong for me to kill the second child to prevent my killing the first.

Of course, consequentialists might argue that moral considerations are unlikely to be equal in the "one on one" case. Not only is there no net gain in lives by killing the second child to prevent one's killing the first, but there is some cost in the effort to do that. Thus, by a utilitarian calculus, it might come out wrong to kill the second child without assigning it any special moral status *vis à vis* the first. But, after all, the cost to oneself of the effort to kill the second child might be balanced either by some slight preference for the first child or by something else. We can imagine examples, then, where the remaining relevant factors, aside from the question of protection, are only whether I kill the second child or allow it to be the case that I have killed the first.

[36] I note a comparable conception of rights in a recent article by Phillip Petit. One feature Petit stipulates is a "non-tradable across persons" feature of rights, explained as "A's right to X is not overridden simply because his [the agent's] violating A's right would minimize the violation of people generally of the right to X on the part of people generally"; see "The Consequentialist Can Recognize Rights," *The Philosophical Quarterly*, XXXVII, 147 (April 1987), p. 46. My conception of a "non-trade-off" feature of rights has some similarity to H. M. Malm's principle that "It is morally impermissible to substitute one person's death for another's (change who lives and who dies) without a good reason for doing so"; "Killing, Letting Die and Simple Conflicts," *Philosophy and Public Affairs*, XVIII, 3 (Summer 1989), p. 248. Malm's principle is weaker than Petit's but at least requires that a good reason for killing cannot be simply that the killing resulted in a situation (impersonally judged) no worse than not killing.

One might also argue that, even without claiming special protection for the second child, there is no reason (particularly if other considerations are balanced) to kill that child to save the first. And, it is said, one needs a reason to act. This argument claims that my somewhat technical sense of 'minimize'—that, whether I kill the second child or not, I have the same (minimum) number of violations—is tricky perhaps, but not to the point. On the other hand (the objection continues), if my violation actually minimized violations in the sense of reducing them, one would in fact have a reason to act. The argument has a point. But it seems to me that saying an action is unreasonable is not equivalent to saying it is forbidden. Consequently, if (when other considerations balance) minimizing right's violations were sufficient reason to violate a right, it still remains permitted, if perhaps unreasonable, to kill the second child to save the first. Moreover, for consequentialists, the mere fact that one brings about equivalent harm by doing *A* or by not doing *A* is not a moral reason for refraining from doing *A*. And, for nonconsequentialists, what makes killing the second child forbidden is not its unreasonableness, but rather the special status of the unthreatened child, a status embodied in the right not to be killed.[37]

It cannot then be correct to say, as David Brink[38] does (explicating Scheffler), that "if a violation of a right is a *bad thing* then it is rational to want to minimize the violation of rights even if we must violate one person's rights in order to do this" (*ibid.*, p. 436; my emphasis). There is an illusion of an intuition here; an illusion that is simply an artifact of the use of evaluative language. The "conjuring trick" has been accomplished if the requirement of reason to minimize is gained by just noting that the violation of a right is a "bad thing." The general concept "bad thing" just will not do that work. The violation of a right is in truth a bad thing, and bad for that reason (that it violates a right). And if Smith violates three particular rights, he perhaps does (to that extent) something worse than Jones who violates only one. But, given the protection of a right, it can remain forbidden to violate one right simply to minimize right's violations.[39]

---

[37] Of course, the first child has exactly the same right. But given the non-trade-off feature of rights, the first child's right is moot in situations where I have to kill another to preserve that right.

[38] "Utilitarian Morality and the Personal Point of View," this JOURNAL, LXXXIII, 8 (August 1986): 417–39.

[39] A more sympathetic interpretation is that Brink is merely claiming that, if the violation of a right is a bad thing, then *ceteris paribus* it is rational to want to minimize right's violations. But this might just be a truism and certainly removes the bite of his point. It leaves open the possibility (one I argue for in the text) that

But suppose we add children in the example above. By killing one child, I could save two, or three, or, in general, N children ($N > 1$) that I have previously threatened. Would that make a difference? This is a difficult but a different question. Some rights can be overridden in some circumstances. But not, as the "one on one" case attempted to show, in order merely to minimize right's violations. Questions remain, of course, about the extent of the protection to those not threatened. I am inclined to think the number of cases of identical harm one would prevent by harming is irrelevant. Others might think that, if one's murder prevented ten identical murders or accidental deaths (but no less), it would be justified. We might say, in general (and admittedly vaguely), violations are permitted if they bring about significantly more good. And although there could be general criteria here, there may be no precise boundary that establishes significance as there is no precise boundary between short and tall.[40]

The special protection due X does entail that I cannot violate X's right to life just to nullify my prior and unjust attempt to kill Y. Admittedly, it does not necessarily follow that it would be wrong to kill X to nullify my prior threat to kill Y and Z. But once we allow some special protection at a moment to those who are significantly unthreatened, there is no Schefflerian "paradox" about an imperative that forbids killing one child to save one, two, or more children from either murder or accidental death. That paradox depended on concluding that, if agency was irrelevant to obligation, then one's causing a violation should be in itself morally irrelevant. I have agreed that agency (at least A1) is irrelevant to obligation but that the protection due the presently unthreatened person is what makes one's causing harm morally relevant. In rejecting the paradox, the door is now open to claiming that a violation must produce significant good rather than simply that it minimize identical violations. Of course, achieving some significant good at the cost of a violation may turn out (in a particular case) also to minimize violations. But it is not constitutive of being a significant good that that occur. The violation

---

violating a right to minimize violations of the same right violates the *ceteris paribus* condition. Brink's use of the category "bad thing" is misleading in another way. If one focuses on welfare, for example, maximizing doctrines like utilitarianism do have some attraction. On the other hand, that attraction is significantly diminished when focusing on something as general as moral disvalue, a category that could include right's violations, failed attempts to harm, unrequited lust, or simply slothfulness. Put simply, there is no moral demand to minimize immorality

[40] For an interesting discussion of how this notion of "significantly more good" might be worked out see Kamm, "The Choice Between People: Common Sense Morality and Doctors," *Bioethics* I, 3 (1987): 256–71

of a right, then, is the property, *P*, which can refer to the consequence of an action and which answers Scheffler's challenge to find some property which both explains, without reference to agency, why a state of affairs is morally objectionable and yet which entails that it can be forbidden to cause one state simply in order to minimize the number of identical states.

An interesting objection to the above argument is that, by making deontology's nonmaximizing character formal (with the non-tradeoff principle), we just make explicit rather than deflate its alleged paradoxical nature. The paradox, some might argue, remains the conflict between this formalism and a more general and better grounded principle of practical reason, i.e., MR.[41] For example, we have admitted that an agent by not committing one violation of a constraint may end up doing something worse, violating two constraints.[42] This appears at odds with the principle of maximizing rationality.

The point would be cogent if, in fact, MR (as Scheffler defines it) requires the interpretation he gives it. Scheffler's definition of MR, however, which ultimately is a principle of instrumental rationality, allows a variety of interpretations from the admittedly vacuous "an agent best achieves the goal of obeying the rules of morality by obeying the rules of morality," to the interpretation that he appears to favor, that, if a state of affairs is morally objectionable and it is morally irrelevant who causes it, then it seems reasonable to cause one state at least to minimize the total number. And there are at least two other interpretations of MR (I can think of) which are not vacuous.

For example, in a moral tradition where what makes an action objectionable is intent (or foresight, or negligence) rather than actual consequences, one can, in fact, minimize what is objectionable by not harming some to save a greater number from one's prior attempts at harm. Those prior attempts constitute wrongs and cannot be undone.[43] If an agent was required to minimize wrongdoing in this sense, then MR is in accord with ordinary deontological constraints. In fact, some commentators on an earlier draft of this paper[44] think this is the simplest way to handle Scheffler's challenge. They contend that the agent does maximize wrongdoing by murdering one child to nullify her prior threat to murder two. The prior threat constituted two violations of a constraint (regardless of whether a death ensued); the murder is a third violation. This view strictly thinks of wrongdoing, including violations of rights, in terms of *mens rea* rather than unjust harm. From this perspective, it is necessarily true that an agent who at any given moment obeys a constraint not to harm some to nullify her own or other's prior threat of harm minimizes wrongdoing overall.

---

[41] This point was made to me by Richard Montgomery and John Whelan

[42] This was a point raised by an editor of this JOURNAL.

[43] I owe the suggestions along these lines as well as other helpful comments to Anthony Ellis.

[44] Both Ellis and David Carr suggested this.

Divorcing intention (negligence, recklessness, etc.) from consequences in the above fashion, and associating wrongdoing only with the former, makes it a matter of course, at least in the trade-off cases discussed, that an agent obeying the constraints of deontology minimizes wrongdoing. So some maximizing demand is met. I have some misgivings, however, about using this argument to show deontology is compatible with MR as Scheffler defines it. MR (for Scheffler) requires a particular goal, but it is not the goal of a deontological agent to minimize her own or anyone else's wrongdoing. Rather, it is simply that, if *mens rea* defines wrongdoing, then an agent obeying a restriction against harming some to aid others will in fact minimize wrongdoing.

Even if we conceived right's violations as requiring actual harm rather than merely attempts at harm, there is, I think, a fourth interpretation of MR (as Scheffler defines it) independent of agency and in accord with ordinary deontological restrictions. One can have the goal, at a given moment, of not violating a right without believing it is of special significance that oneself as opposed to others not violate the right. And the agent can quite rationally believe, given the special protection due the unthreatened, that violations are forbidden just to minimize instances of identical violations. Scheffler apparently believes that, if an agent prefers no right's violations to one, she should prefer less violations to more, unless what is significant for her is that she not violate a right.[45] But this ignores the possibility that resistance to the goal of minimizing total violations (the goal he believes required by MR) follows from the nature of rights rather than from special concern with one's agency.

It seems to me, then, that MR—choosing the best means for a desired goal—does not entail, without begging the question in favor of consequentialism, any demand to minimize my own or others' right's violations. The goal in the conception I have presented is temporally indexed to respect at any given moment the right of one who is innocent and unthreatened. An agent best meets the goal by respecting the right at that moment, though as a result she may end up having done something worse overall, for example, having violated the rights of two people. We can even say that the agent does something worse at the moment by violating the third person's right than by not violating it. In this sense an agent, by not violating a right at a given moment, does minimize her own wrongdoing at that moment. But, again, this is not because of some special moral significance to her agency, but rather because of a temporally indexed goal that reflects the right of a patient. And it is precisely that index which allows us to resist the notion that, implicit in a commitment not to violate a right, is an agent's commitment to reduce her own or others' right's violations overall. Thus, an agent can both *do* something worse at a moment by violating a right than by not violating it and end up *having done* something worse by not violating the right. That is, the agent, by not violating one right at time *t*, can end up having violated two rights. But at *t* she does not violate two rights, but allows that to occur. There would only be a paradox here if (as Scheffler appears to believe) implicit in a conception of wrongdoing is either an imperative of practical reason to minimize wrongdoing overall, or (as he thinks true, but irrational, about deontology) a demand on an agent to minimize her own wrongdoing. I have suggested a third alternative, a temporally indexed commitment that reflects the right of an unthreatened patient.

---

[45] See, again, "Agent-Centered Restrictions, Rationality, and the Virtues," p 415.

FURTHER COMMENTS ABOUT AGENCY, DOING VERSUS ALLOWING, AND DOUBLE EFFECT

As for A3, the temporally indexed version of the doing-allowing distinction, it retains only derivative significance in a patient-based defense of deontology.[46] Violating constraints against harming some to aid others requires intentionally or knowingly (or negligently) at a given moment altering the world to threaten harm to one then significantly unthreatened. Therefore, violating a constraint at a moment is *to do* something rather than (at that moment) *merely allowing* it to occur. Of course, even merely allowing something, *X*, to occur may invariably entail that one does something else, *Y*. In allowing the general to drown, I have done something, altered the course of the war, although I have not necessarily killed the general. But the contrast I want is between doing *X* (by commission or omission) at a moment and merely allowing at the moment more instances of *X* to occur. That contrast (between doing and allowing at a moment) is needed for the puzzle cases about harming some to prevent the same

[46] That it must remain significant was pointed out to me by Ellis.

harm to others even to get started, and is based on whether before one's action (or inaction) another is significantly threatened with *X* happening to him.[47]

The temporal indexing in A3 is required to support formally the strong belief that it can be wrong at a moment for an agent to cause harm to nullify her own prior attempts at more harm. Without the indexing, the agent (in single-agent cases) is faced simply with a (temporally unrestricted) choice of how many people ultimately to harm. But the temporal indexing is neither fundamental nor an arbitrary device to "save the (moral) phenomena." It is based on the special protection due those who (at the moment) are both innocent and unthreatened.

One objection to the above comments is that one gains little by calling my conception a patient-based deontology, since in the cases considered there is a "reciprocal entailment" between the duties of agents expressed by A3 and the rights of patients.[48] Why not allow A3, then, to be fundamental, and deontology to be agent- rather than patient-based. The alleged "special protection" due the innocent merely reflects, on this account, the fundamental duties of agents at any given moment not to bring the innocent under threat.

The answer, I believe, is that, though there is a reciprocal entailment of rights and duties in the cases I have discussed, the order of explanation is asymmetrical. It makes sense to explain the temporally indexed duty in terms of the special protection due someone who is unthreatened, and not the other way around. That right is supervenient on the status at a moment of an innocent person's pursuing her unthreatening interests. This is the default position. It is the moral significance of that status that accounts for the fact that the agent's duty must be temporally indexed. Otherwise, we would wonder why a person especially concerned with her own agency would not wish to minimize her own wrongdoing overall (i.e., in a temporally unrestricted way). The temporal index reflects the moral status of the unthreatened person whose life the agent might (but should not) trade to save others she has previously threatened. The order of explanation is from this status to the duties of agents. It is in this sense that I consider deontology patient-based. The distinction I wish to draw, again, is between this view of deontology and one which Scheffler seems to espouse, which conceives an agent concerned ultimately with her own doings as opposed to overall consequences.

---

[47] Omissions as well as commissions can certainly count as "doing *X*." For example, starving one's child by simply omitting to feed it is killing it. The deeper issue is what counts as a normal state of the world such that an omission leading to a death is a killing. Special duties and related expectations are not sufficient. The lifeguard who fails an obligation to make an easy rescue has not killed the swimmer. The omission must count for us as the initiation of the threat of death

[48] This point was suggested by an editor of this JOURNAL.

If I am right about this "privilege of innocence," i.e., the special protection due those who are unthreatened and unthreatening, then double effect's distinction between intention and mere foresight lacks ultimate significance. This conclusion is unsettling since it is in fact double effect (or some closely related distinction) that appears to account for the quite powerful thought that what is particularly wrong is to harm some *in order* to aid others, as opposed to harming some as a collateral but unintended effect of aiding others. Discounting the distinction made by double effect is to discount what is intuitively and historically significant in deontology. One might attempt to incorporate double effect in a patient-based deontology by contending that the privilege of innocence is embodied in a right not to be intentionally harmed (or in the Kantian variant of double effect, a right not to be used merely as a means). There would be no comparable right not to be foreseeably harmed (or knowingly harmed in a way where one was not used). My belief now is that such a view would be patient-based only in a secondary sense. The primary wrong would be "aiming" in a certain way at the harm one caused. The right of someone not to be harmed in that way would be derivative.[49]

SOME CONCLUDING REMARKS

My aim was to sketch a patient-based defense of deontology to answer Scheffler's challenge that any defense will violate the apparently plausible goal that, other things being equal, one should act so as to minimize what (in an impersonal sense) is morally objectionable. The defense is based on giving moral weight to someone's status at a moment of not being significantly threatened with harm. We embody this recognition by ascribing rights, for example, the right not to be killed. What is salient about rights is the "non-trade-off principle," i.e., minimizing right's violations is not a sufficient reason for violating a right. The principle follows directly from recognizing the special protection due the person not threatened. That protection entails that it would be wrong for me to kill *X* to nullify my prior attempt to kill *Y*. The extent of this protection is admittedly unsettled, but at least we can argue that there is no paradox of the sort Scheffler raises about a restriction against killing one person though that would save even two or more others.

---

[49] Nagel, defending a strong prohibition against intentional harm, mentions the "special outrage" the victim feels in being "deliberately harmed" for the ("even greater") good of others. This suggests he thinks the right involved is the right not to be harmed as an instrument to achieve another's ends. There would be no comparable right not to be unintentionally (though foreseeably) harmed. See *The View From Nowhere*, p. 184. Yet one can develop examples, consistent with the strictures of double effect, where concern for the victim vanishes entirely as a consideration. Imagine a case where, given the good to be achieved, I believe (as a subscriber to double effect) that it is permitted to foreseeably, painfully, but unintentionally kill Smith. Smith requests that, instead, I intentionally and painlessly kill him (a possibility he discovered) in order to achieve the same good. As a subscriber to double effect, I refuse, and kill him painfully. The thought experiment, though admittedly strange, suggests that double effect's conception of morality as the way one's will relates to what one does is perhaps not at bottom primarily concerned with either welfare or rights It may be the only truly agent-centered morality.

And, in addition, I noted that this defense of deontology is in accord with the "principle of maximizing rationality," MR, given Scheffler's quite general formulation of that principle; for an agent can have the goal at each moment of honoring the special protection embodied in the right of a person not to be harmed. And she best achieves the goal by not harming the person. Reason is satisfied.

Of course, no defense of rights, whether Kantian or contractarian or whatever, may prove successful. In particular, the notion I have sketched of "inertia" in the moral domain may not pan out. But the strength of the belief that it is wrong to cause harm even to minimize one's own harming suggests that moral concern is with the person unthreatened at that moment and not with whether we, rather than others (or natural forces), cause the harm. It is that concern which motivates the ascription of rights. I would note, as well, and somewhat in passing, that nothing in Scheffler's challenge appears to disallow the ascription of rights and therefore of right's violations. His point, even here, is that, if the violation of a right is objectionable regardless of who causes it, it seems paradoxical to forbid a violation whose sole effect is to minimize violations. But this raises the specter of paradox at a much too abstract and therefore misleading level. If a conception of rights is defensible, there is nothing unreasonable about a patient-based moral point of view which denies that, implicit in thinking right's violations to be objectionable, is a demand to minimize violations overall.

RICHARD BROOK

Bloomsburg University

BOOK REVIEWS

Studies in the Way of Words. PAUL GRICE. Cambridge: Harvard University Press, 1989. 394 p. Cloth $32.50.

This collection of essays brings together Paul Grice's writings on the problems of meaning broadly conceived. Ten of them have previously appeared in print; ten have not. His William James Lectures, presented in 1967 and revised in 1987, are here published in their entirety for the first time.

In various places, Grice indicates how his thoughts about problems concerning meaning developed. He describes with warmth and nostalgia his participation in the movement known as ordinary-language philosophy (OLP) which arose at Oxford after World War II. He further tells us that he never gave up the idea that ordinary language is both a suitable topic for philosophical investigation and a genuine source of philosophical insight. But even if Grice was one of the chief and most gifted practitioners of OLP, he was also, perhaps, its most penetrating critic. Following the maxim that the meaning of an expression is its use in the language, it was common among OL philosophers to point out (almost always correctly) that some philosopher had employed an expression in a way in which it is not ordinarily employed. For example, J. L. Austin would note such things as that in a dense fog it might be natural and correct to say that it looks like there is a pig over there or, perhaps, there appears to be a pig over there, but on a perfectly clear day, with the pig plainly in view, it would surely be inappropriate, odd, a misuse of language to speak in either of these ways. This and similar remarks about how we actually employ certain expressions were supposed to reveal fundamental flaws in various traditional theories of perception, in particular, those which attempted to ground empirical judgments in judgments about appearances.

Grice smelled a rat. In a presentation before the Aristotelian Society (1961), he undertook a defense of a causal theory of perception, at least against criticisms of this kind. He maintained that conversational oddness *by itself* did not show that an utterance is either meaningless or false. He pointed out that conversational oddness can arise for various reasons, one being that the speaker's use of an expression may *suggest* or *imply* something false or misleading. This, Grice noted, is compatible with the truth and, of course, the meaningfulness of what the speaker has actual said. He then went on to show in close detail that a causal theory of perception could be defended against OLP criticisms along just these lines.

0022-362X/91/8804/213–9

This contrast between what a speaker *actually says* and what the saying of it *conversationally implicates* (a label Grice introduced later), once pointed out, seems immediately plausible. In fact, Grice could have formulated his defense of the causal theory of perception quite effectively without going beyond this intuitive level. It was, however, completely uncharacteristic of Grice to let matters remain at an intuitive level. Since his argument depended upon a contrast between what the speaker actually says (the literal meaning) and what the saying of it implies (its conversational implicatures), he felt obliged to provide a detailed theoretical account for each of them. Both of these tasks were pursued in Grice's William James Lectures.

"Logic and Conversation" (Essay 2) was an underground classic before it first appeared in print in 1975. In this essay, Grice developed his account of conversational implicature, basing it on the idea that linguistic exchanges are instances of shared rational purposive activity governed (for the most part, at least) by what he called the cooperative principle and a set of more specific maxims that fall under it. Although there is room for disagreement concerning details, Grice's general approach to these pragmatic features of language is almost certainly correct, since it is able to give plausible and detailed accounts of indirect uses of language over a very wide range of cases. This part of his program is money in the philosophical bank.

As noted, Grice's attention was first drawn to the phenomenon of conversational implicature in his efforts to defend a causal theory of perception against a style of criticism common among OL philosophers. In fact, Grice constantly set himself apart from other OL philosophers by defending at least the intelligibility of many traditional philosophical positions that OL philosophers were particularly interested in rejecting. In various places in his works, he invokes the notion of conversational implicature to defend a causal theory of knowing (Essay 3), a sophisticated version of a causal theory of meaning (throughout), something like a correspondence theory of truth (Essay 3), and Bertrand Russell's theory of definite descriptions (Essay 17). It seemed to be Grice's particular calling to resuscitate philosophical positions that OLP had claimed to have buried.

In the third of his William James lectures, "Indicative Conditionals," Grice employs this method of philosophical resuscitation in an effort to show, among other things, that the logician's notion of material implication specifies the correct semantic content of ordinary 'If *p*, *q*' constructions. An important limitation on this discussion is that Grice does not take this approach to 'If *p*, *then q*', which he takes to mean 'Supposing *p*, (then) *q*' (63). Excluding such indicative conditionals that, as he thinks, express reasoning under a supposition, Grice invokes the notion of conversational implicature to defend the material implication account of indicative conditionals.

Grice's defense is most effective in dealing with problems that turn upon the so-called paradoxes of implication. "Surely it is an abuse of language," the OL philosopher will say, "to employ an indicative conditional just because its consequent is true or its antecedent false." Grice acknowledges that it would indeed be odd to use an indicative conditional when one knows either of these things, but this oddness, he argues, arises from using the weak construction 'If *p*, *q*' when we are in a position to use the stronger constructions 'not *p*' or '*q*'. Grice calls this a violation of the rule of quantity. It is this violation of the rule of quantity which accounts for the conversational oddity of employing an indicative conditional in such a context, and the oddity, thus explained, is compatible with the truth and meaningfulness of what has been said. This seems right.

This, however, is not the end of the story, for Grice encountered deeper difficulties in dealing with the denials of indicative conditionals. As Charles L. Stevenson noted some time ago, if we treat indicative conditionals as material conditionals, we can pass from the (apparent) truth 'It is not the case that if God exists, there is evil in the world' to the conclusion that God exists and there is no evil in the world. A successful theology and a theodicy should be harder work. Grice attempts to deal with the problems that arise from the denials of conditionals by introducing a bracketing device intended to control the scope of the negation sign. Sometimes the force of denying a conditional is given by the denial of the material conditional and here no component is bracketed. Grice gives clear and unforced examples of conditionals of this kind. Sometimes the antecedent is protected from negation by brackets—perhaps because it is common ground between the parties in the conversational exchange—and then the denial attaches only to the consequent. At other times, the conditional is not denied at all, and the denial is aimed at some conversational implicature of the use of the conditional. In effect, the entire conditional is bracketed, though Grice did not exactly put it this way (see 81–2).

Grice's strategy, then, for defending the thesis that the semantic content of 'If *p*, *q*' constructions is given in the counterpart material conditional was to use, in tandem, the notions of conversational implicature and bracketing. Later (in Essay 17) Grice uses the same double strategy in defending Russell's theory of definite descriptions against criticisms leveled against it by Peter Strawson. It must be said that Grice does not work out his theory of bracketing in full detail, and it remains to be seen whether it can be made to work.

The first four William James Lectures were primarily dedicated to developing the theory of conversational implicature and then applying it to certain philosophical issues. The second main task of his William James Lectures was to develop a theory that would explain speaker's or utterer's meaning using "the notion of intention in a nested kind of way" (283). Grice's interest in this program in fact antedated his concern with conversational implicature. "Meaning," where this theory was first sketched, was published in 1957, and apparently given as a lecture almost ten years earlier in 1948. About this aspect of his philosophizing, Grice remarks that "it has given me more trouble, and it has also engendered more heat" (349) than any other.

In the early stages of developing his position, Grice drew a distinction between natural meaning and non-natural meaning—the difference, for example, between smoke meaning fire and John meaning his uncle when he uses the expression 'the old codger'. (Here other philosophers have commonly drawn a distinction between natural signs and conventional signs, but, for reasons I shall examine in a moment, Grice did not want to draw the distinction in this way.) He presented what we might call his core account of non-natural meaning (meaning$_{NN}$) in these words:

> "A meant$_{NN}$ something by x" is roughly equivalent to "A uttered x with the intention of inducing a belief by means of the recognition of this intention" (219).

Starting with this relatively simple specification of meaning$_{NN}$, Grice's position passed through an extraordinary series of refinements and elaborations. Some of these elaborations were the result of technical demands, for example, to show that this simple formula could be extended to deal with the full variety of speech acts as they arise in various linguistic settings. I shall not examine such details here. The second pressure for elaboration came from the need to respond to counterexamples. Here there is only space to say a few programmatic things about these counterexamples.

Schematically, the original counterexamples to Grice's core account of meaning$_{NN}$ seemed to work in the following way: if the Gricean definition of meaning$_{NN}$ involves the mutual recognition of intentions up to some level L, then it always seems possible to produce a counterexample to the definition by introducing a bad or sneaky intention at level $L + 1$. In Grice's words, "the hearer is represented, at some level or other of embedding, as having . . . a misapprehension with regard to what is expected of him" (302–3). A quick way of dealing with counterexamples of this kind would be to add a negative clause to the definition proscribing the use of such sneaky intentions. In a sense, Grice does take this way out of his difficulty, but only after addressing a much deeper difficulty.

Setting aside problems with particular counterexamples, in "Meaning Revisited" (Essay 18), Grice argues that his account of meaning$_{NN}$ does, in fact, commit him to saying that, in fully or optimally meaning something, one is committed to the unlimited nesting of intentions. Then, instead of attempting to revise his position to avoid this consequence, he actually embraces it, thus giving his position a remarkable new turn. Grice arrived at this position from the following line of reflection:

> One might start with the idea that when some speaker S utters some sentence to a hearer A, meaning by it that p, he does this wanting A to think 'p'. That is, at stage one we have "S wants A to think 'p,' " where p represents the content of A's thought or intended thought. However, for reasons that came up long ago, having to do with the distinction between natural and nonnatural meaning, we cannot stop at stage one. We have to proceed to stage two, at which we get "S wants A to think 'p, on the strength of the idea that S wants A to think "p" ' "; and so on (300).

This is obscure, but, very roughly, *S*'s intention that his intention be recognized cannot itself be conveyed merely as a natural sign. *S*'s intention that his intention be recognized is also something that *S* intends to be recognized. It seems, then, that speakers are always in need of one more intention in order to mean fully what they intend to mean. As Grice puts it:

> what we specify A as intended to think of as the reason for thinking "p" is always one stage behind what the speaker envisages as the reason why he wants A to think "p" (300).

It follows from this that the conditions necessary for someone to mean$_{NN}$ something are logically impossible to fulfill (see 301). It seems, then, that, on Grice's account of meaning, meaning something is something a speaker is never able to pull off.

Grice's radical proposal for dealing with this difficulty is to suggest that on certain occasions "it is legitimate, or perhaps even mandatory, for us to ***deem*** [the speaker] to satisfy this unfulfillable condition [of endlessly nested intentions]" (emphasis added, 302). But what is deeming? And why countenance it? Grice does not answer these questions in detail. He suggests an interesting parallel with a problem in the theory of knowledge. If to know something, you cannot be wrong about it, then, according to Grice, it will be correct to say that "*strictly speaking* the only things that can be known are necessary truths" (301). He then adds:

> However, that does not restrict us to supposing that people who talk about knowing other things are using the word "know" improperly: all it requires is that there should be some license to apply the word non-strictly to things which in some way approach or approximate to the ideal cases (301–2).

With this suggestion, the burden of the argument shifts to the evaluative notion of an *ideal* case. In fact, in this volume, Grice says little about his views concerning values except for indicating, in various places, his commitment to the existence of objective values and stating, very darkly, that an appeal to objective values will be necessary "to exhibit the anatomy of speaker meaning itself" (299). This presumably is related to specifying the conditions under which it is legitimate or even mandatory to deem a condition fulfilled when, in fact, we know it to be unfulfillable. These reflections also provide a rational ground for ruling out counterexamples that involve sneaky intentions. Such sneakiness in intentions provides us with a positive reason for not deeming them ideal.

Grice's reflections on these matters are presented in a section of "Meaning Revisited" candidly entitled "The Mystery Package." The package will remain a mystery until Grice's unpublished writings on values and teleology find their way into print. Members of the Berkelischer Kreis indicate that progress is being made on this front.

I have said nothing about the final step in Grice's account of meaning: the attempt to explain word meaning and sentence meaning in terms of utterer's meaning. Roughly: (1) utterer's meaning is to be explained in psychological terms, primarily, with reference to nested intentions, and (2) sentence meaning and word meaning are to be explained in terms of utterer's meaning. Thus, in the end, complex word and sentence meanings are grounded in complex psychological states. One difficulty with this view is that certain complex intentions, as Wittgenstein has claimed, seem available only to creatures who *already* possess a language. A cat may stalk a bird with the intention of catching it, but it seems that a cat could not intend to stalk a bird a week from Tuesday. The later intention, it seems, could only be formed by a creature already in possession of a language. Similarly, the complex nested intentions that Grice invokes to explain utterer's meaning would seem to be unavailable to a creature who did not already possess a language.

I do not think that these considerations provide a knock-down argument against Grice's account of utterer's meaning and its relationship to word and sentence meaning. It might be possible to tell a reasonable story showing how prelinguistic intentions (of the kind that animals seem to have) can somehow undergird a system of primitive conventions that, through bootstrapping, can develop into the rich system of linguistic conventions that we actually use in saying things meaningful. Grice actually tells a story something like this in "Meaning Revisited," but unfortunately, right from the start, the intentions that Grice invokes seem to be those which only someone who already commands a conventional language could form. Thus, since he has not shown how semantic bootstrapping can actually take place, Grice's plausibility argument for his general program seems not to have been made out.

I have concentrated on two aspects in Grice's philosophical career: (i) his development of a theory of conversational implicature grounded in a rich conception of language as a rational, cooperative, goal-directed activity, and (ii) his attempt to develop a general theory of meaning based on the fundamental notion of intentions taken in a "nested kind of way." It seems to me that the first program has been largely successful and has had a fundamental impact upon the way philosophy has been practiced in the second half of this century. The second program has been less successful and may, in the end, seem wrongly conceived. But both programs reveal a mind constantly trying to get to the bottom of things. Paul Grice was a philosopher incapable of superficiality.

ROBERT J. FOGELIN

Dartmouth College

Against Method (revised edition). PAUL FEYERABEND. New York: Verso, 1988. vii, 296 p. Cloth $50.00, paper $16.95.
Farewell to Reason. PAUL FEYERABEND. New York: Verso, 1987. 327 p. Cloth $49.95, paper $15.95.

Unlike most revised editions, the new *Against Method* should be noticed. Most of the original[1] is still there, although if you scanned

---

[1] *Against Method* (London: New Left Books, 1975) The true originals of the book are more like "Problems of Empiricism," in *Beyond the Edge of Certainty*, R.

0022-362X/91/8804/219–23

only the indices of the two editions you might not guess that they are versions of the same book. Ephemeral polemics with minor figures have been removed, and the showpiece confrontation with Imre Lakatos's "Methodology of Scientific Research Programmes" has been dissolved. The show-off catch phrases that everyone remembers—anarchy/dada, "anything goes"—have been demoted, and a solemn introduction written for Chinese makes plain what the casual reader of the first edition often missed, namely, what the book was about. "*The events, procedures and results that constitute the sciences have no common structure*"; far-reaching changes of outlook in the sciences have happened in different ways, and given different impulses to future research (1). The book has an argument by powerful examples that make its central thesis plausible.

The new last chapter is an engaging intellectual autobiography with an important reminder of some of the people who (in Feyerabend's self-conception) profoundly affected his thinking: G. E. M. Anscombe was a primary source, as were Philipp Franck, Victor Kraft, and Felix Ehrenhaft. And it was C. F. von Weizsäcker who put a stop to Feyerabend's alienated assaults on quantum physics. The autobiography ends with an honorable confession of regret. He once rather bombastically thought that his relativistic philosophy could serve the students that Berkeley brought in from the ghettos and barrios of California. Now: "it is conceited to assume that one has solutions for people whose lives one does not share and whose problems one does not know. . . . My own view . . . was still a view, an abstract fancy I had invented and now tried to sell without having shared even an ounce of the lives of the receivers. This I now regard as insufferable conceit."

The heading for the autobiography states that, above all, the motive force behind Feyerabend's work is anger: anger at intellectual conceit, at "treacly phrases" of intellectuals, at destruction of, interference in, or control of other people's lives. As for those who pro-

---

Colodny, ed. (Englewood Cliffs, NJ: Prentice-Hall, 1965), pp. 145–260; and "Problems of Empiricism Part II," in *The Nature and Function of Scientific Theories*, R. Colodny, ed. (Pittsburgh: University Press, 1971), pp. 275–354; neither of which occurs in *Problems of Empiricism: Philosophical Papers*, Vol. 2 (New York: Cambridge, 1981), but both of which are largely retained in the revised edition of *Against Method*. There is also a German version whose third edition is different from any English one, and which embodies the final, eponymous, chapter of *Farewell to Reason*; cf. *Wider den Methodenzwang* (Frankfurt: Suhrkamp, 1986) claim or fear that there is "cultural chaos" that "pervades the essence of our culture," they are speaking about the rantings of a handful of highbrows. We should rather fear what François Jacob calls the threat of "monotony and dullness."

Feyerabend's considered philosophy is strictly critical, constructive only in urging people to learn from each other but otherwise to live and let live. It rejects without sadness all claims to permanence or foundations. His arguments are ad hominem, and his metaphilosophy is a commitment to ad hominem criticism. So he can cheerfully tell the mandarins who bewail "cultural chaos" to stuff it and watch the Superbowl and *Dynasty* to see how little chaos there is in their culture—and then on other pages regret how one technology and its values can sweep aside all others, empowering *Dallas* and (Association) football not only in Rome but also in Burkino Faso. He undoubtedly thinks that relativism is some sort of prophylactic on both fronts.

The longest essay in *Farewell to Reason,* "Notes on Relativism," does not urge relativism as a single monolithic doctrine. Instead, a series of eleven loosely relativist positions is rehearsed. All are presented as reactions against the militant claims of capital *R* Reason. He has no objection to being reasonable or objective, by whatever immediate canons we may have to hand. He is against the notion that there is some ultimate objective and rational way of finding out what is true. "The ideas of truth, reality and rationality make excellent practical sense"; Feyerabend opposes only the metaphysical constructions therefrom, the One Truth, the One Reality, the One Rational Method.

The eleven subsequent essays in the collection vary in interest. They are best when they present a simple figure as complex, as in the discussion of Ernst Mach, and worst when they present a complex figure as simple, as in the essay on Karl Popper. The deflating of Popper is irrelevant to American culture, where Popper is of no significance, but elsewhere he is the most widely read philosopher of the twentieth century. Hence the placing of Popper within the problematic of logical positivism may be of value, and there is a nice diagnosis of why so many scientists took him so seriously. There is a very tidy essay on Aristotle's theory of mathematics, and a brisk, accurate, rebuttal of Hilary Putnam's flippant contention that Feyerabend's idea of incommensurability and relativism are self-refuting. The concluding chapter, called "Farewell to Reason," is based on a reply to two German volumes of essays about Feyerabend. Like most replies of the "Schilpp" sort, that reply did not come off well in the first place; it is not satisfying on its own as the conclusion to this book.[2]

Undoubtedly, these essays will be read chiefly to find out Feyerabend's latest thoughts about relativism. I believe he says one thing that is completely wrong, one that is happily belied by his own practice. He says that the kind of relativism he presents is "about human relations. It deals with problems that arise when different cultures, or individuals with different habits and tastes, collide." But he is not talking about the collision of cultures in general. Despite the occasional third-hand references to exotic peoples, his relativism is "ours" and is our Greek heritage. Look at the proper names in the index. Bohr, Einstein, Galileo, Mach, and Popper have big entries, but after that, it is Greeks all the way. That is as it should be. Only "Westerners" have a problem about relativism, and we know where it comes from. *Farewell to Reason* may be the closest we will get to a book Feyerabend has never quite written, an assault on the entire tradition engendered by those imperial Greek rationalists who had us replace magic by mathematics.

Relativism (as opposed to what Feyerabend rightly calls opportunism—learn from the others what is handy) does not arise automatically from cultural confrontation. Peoples have been crossing paths since the beginning, with consequences that range from ruthless genocide, through trade-offs and the division of labor, to straightforward indifference. Most histories, be they of the Pacific Northwest of this continent or 3000 years of Persia, have the usual mess of slavery, pillage, marauders, and mayhem interspersed with quite a lot of more peaceful commerce. Just like "our" history. Many peoples, and certainly Westerners, have been confident of their own ways, often to the point of extinguishing others. But it is only "we" who have been relativists, because of our Greeks who articulated the abstract concepts of truth, objectivity, and reason.

Travelers have always known about and been amused or disgusted by the pollution rules of other people, but it was Herodotus who (on Feyerabend's undisputed telling) first drew the relativist conclusion. Darius set up Herodotus by introducing Greeks to people who ate their dead parents, and who thought that the Greek practice of cremation was heinous. Darius was merely (I suppose) mocking the

---

[2] H. P. Duerr, ed., *Versuchungen*, 2 vols. (Frankfurt: Suhrkamp, 1980, 1981). An English translation of the essays and reply is being prepared by G. Munévar, *Beyond Reason: Essays on the Philosophy of Paul Feyerabend* (forthcoming in the Boston Studies in the Philosophy of Science Series).

Greek objectivity mongerers, but Herodotus was fully serious in the way that only a Western rationalist can be. What if my enthusiasm for science is no less a matter of custom than my revulsion at the thought of eating my dad's corpse? This tale (which Feyerabend has used for decades) stands for the intense Greek debates about objectivity and relativism which their successors have sometimes muted but never forgotten. And it was not ethical relativism that came first, at least on the testimony of Aristotle, but qualms about this new conception, objective knowledge.

To repeat, relativism is not about "what happens when different cultures . . . collide." It is about what happens when "our" apparently unique concept of objectivity meets anyone else. Just as objectivity was located not in things but in what lay behind things, or in the reasons traversed and surveyed by the mind, so the doubts that follow the examination of other knowledges thrust us inward upon ourselves. "Anything goes," Feyerabend confessed in the preface to the new *Against Method,* was "the terrified exclamation of a rationalist who takes a closer look at history." And the eleven relativist theses of "Notes on Relativism" are the rationalist whimpers that follow this cry of terror.

IAN HACKING

University of Toronto

NEW BOOKS: ANTHOLOGIES

ADAMS, PARVEEN, and ELIZABETH COWIE, eds.: *The Woman in Question:* m/f. Cambridge: MIT, 1990. 383 p. Cloth $25.00.

BARRETT, ROBERT B., and ROGER F. GIBSON, eds.: *Perspectives on Quine.* Cambridge: Blackwell, 1990. xxiii, 358 p. Cloth $45.00.

BRICKER, PHILLIP, and R. I. G. HUGHES, eds.: *Philosophical Perspectives on Newtonian Science.* Cambridge: MIT, 1990. vii, 248 p. Cloth $34.95.

COHEN, PHILIP R., JERRY MORGAN, and MARTHA E. POLLACK, eds.: *Intentions in Communication.* Cambridge: MIT, 1990. xi, 508 p. Cloth $45.00.

EVERSON, STEPHEN, ed.: *Companions to Ancient Thought 1: Epistemology.* New York: Cambridge, 1990. viii, 288 p. Cloth $54.50, paper $16.95.

FINE, ARTHUR, MICKEY FORBES, and LINDA WESSELS, eds.: *PSA 1990, Volume One.* East Lansing, MI: Philosophy of Science Association, 1990. xxix, 599 p. Cloth $22.00, paper $20.00.

FRENCH, PETER A., THEODORE E. UEHLING, JR., and HOWARD K. WETTSTEIN, eds.: *Midwest Studies in Philosophy: Volume XV: The Philosophy of the Human Sciences.* Notre Dame: University Press, 1990. 464 p. Cloth $49.95, paper $21.95.

GEORGE, ALEXANDER, ed.: *Reflections on Chomsky*. Cambridge: Blackwell, 1990. xii, 271 p. Cloth $45.00.

GILL, CHRISTOPHER, ed.: *The Person and the Human Mind: Issues in Ancient and Modern Philosophy*. New York: Oxford, 1990. x, 284 p. Cloth $55.00.

GOLDBERG, DAVID THEO, ed.: *Anatomy of Racism*. Minneapolis: Minnesota UP, 1990. xxiii, 357 p. Cloth $39.50, paper $14.95.

GOODIN, ROBERT E., and ANDREW REEVE, eds.: *Liberal Neutrality*. New York: Routledge, 1990. iv, 219 p. Cloth $49.95.

JACOBUS, MARY, EVELYN FOX KELLER, and SALLY SHUTTLEWORTH, eds.: *Body/Politics: Women and the Discourses of Science*. New York: Routledge, 1990. vii, 107 p. Cloth $39.50, paper $13.95.

KELLY, MICHAEL, ed.: *Hermeneutics and Critical Theory in Ethics and Politics*. Cambridge: MIT, 1990. xi + 285 p. Paper $13.95.

KEMP, T. PETER, and DAVID RASMUSSEN, eds.: *The Narrative Path: The Later Works of Paul Ricoeur*. Cambridge: MIT, 1990. viii, 121 p. Paper $10.95.

KOELB, CLAYTON, ed.: *Nietzsche as Postmodernist: Essays Pro and Contra*. Albany: SUNY, 1990. vii, 350 p. Cloth $59.50, paper $19.95.

KYBURG, HENRY E., JR. et al., eds.: *Knowledge, Representation, and Defeasible Reasoning*. Boston: Kluwer, 1990. xxii, 423 p. Cloth $99.00.

LYCAN, WILLIAM G., ed.: *Mind and Cognition: A Reader*. Cambridge: Blackwell, 1990. ix, 683 p. Paper $19.95.

MACDONALD, SCOTT, ed.: *Being and Goodness: The Concept of the Good in Metaphysics and Philosophical Theology*. Ithaca, NY: Cornell, 1991. ix, 328 p. Cloth $43.95, paper $14.95.

OSHERSON, DANIEL N., et al., eds.: *An Invitation to Cognitive Science: Volumes 1–3*. Cambridge: MIT, 1990. xix, 273 p.; x, 356 p.; x, 308 p. Each cloth $37.50, paper $18.95.

PETERSON, GRETHE B., ed.: *The Tanner Lectures on Human Values, Volume XI, 1990*. Salt Lake City: Utah UP, 1990. ix, 578 p. Cloth $25.00.

SAVAGE, C. WADE, ed.: *Scientific Theories*. Minneapolis: Minnesota UP, 1990. ix, 429 p. Cloth $30.00.

SCHNEEWIND, J. B., ed.: *Moral Philosophy from Montaigne to Kant: An Anthology: Volume I and Volume II*. New York: Cambridge, 1990. xxi, 664 p. Vol. I: Cloth $49.50, paper $14.95. Vol. II: Cloth $49.50, paper $13.95.

SEREQUEBERHAN, TSENAY, ed.: *African Philosophy: The Essential Readings*. New York: Paragon, 1991. xxii, 250 p. Paper $13.95.

SOLOMON, ROBERT, and KATHLEEN HIGGINS, eds.: *Reading Nietzsche*. New York: Oxford, 1990. xiv, 251 p. Cloth $32.50, paper $12.95.

SORABJI, RICHARD, ed.: *Aristotle Transformed: The Ancient Commentators and their Influence*. Ithaca: Cornell, 1990. x, 545 p. Cloth $65.00.

WOLTER, ALLAN B.: *The Philosophical Theology of John Duns Scotus*. MARILYN MCCORD ADAMS, ed.: Ithaca: Cornell, 1990. ix, 356 p. Cloth $47.50.

WREN, THOMAS E., ed.: *The Moral Domain: Essays in the Ongoing Discussion between Philosophy and the Social Sciences*. Cambridge: MIT, 1990. xxix, 414 p. Cloth $35.00.

YOURGRAU, PALLE, ed.: *Demonstratives*. New York: Oxford, 1990. vi, 275 p. Cloth $39.95, paper $12.95.

Leibniz's Monadology
An Edition for Students

Nicholas Rescher

A collection of materials that constitute a commentary on the *Monadology* by Leibniz himself.

"Well conceived, well executed, and providing an extremely useful tool, one which has been sorely needed. . . . This book fulfills a critical need: Leibniz's major work has never before been readily accessible for students. In Rescher's presentation it will be."
—George Gale, University of Missouri, Kansas City

$39.95 cloth / ISBN 0-8229-3670-4
$19.95 paper / ISBN 0-8229-5449-4

University of Pittsburgh Press
c/o CUP Services, Box 6525, Ithaca, NY 14850 • 800-666-2211

The British Journal for the Philosophy of Science

Editor: G.M.K. Hunt, University of Warwick

Forthcoming articles will include:
Experimental Evidence and Psychotherapy
Cognitive Neuropsychology and the Philosophy of Mind
Factor Analysis and Objectivity
Agency and Probabilistic Causality
Simplicity and Falsifiability
Statistical Regression
Defining Chaos
Mental Content in Animals
Creativity and Science
Is Classical Rationality always adaptive?

Subscription rates Volume 41, 1991 (4 Issues)
£30.00 UK & Europe; US$62.00 Elsewhere
Individuals may join the British Society for the Philosophy of Science and receive the journal at reduced prices

To subscribe, obtain further information or for a *free* sample copy of this journal, please write to: Journals Marketing Department (X), Oxford University Press, Southfield Road, Eynsham, Oxford OX8 1JJ, UK

New for Spring

Instrumental Realism

The Interface between Philosophy of Science and Philosophy of Technology

By Don Ihde

"Ihde is perhaps uniquely situated to provide authoritative accounts of such diverse philosophical traditions as those involved in current explorations of the technology of scientific instruments. . . Ihde's book breaks new ground and . . makes an important debate accessible."—Robert Ackermann

The Indiana Series in the Philosophy of Technology

cloth $29 95

paper $12 50

Re-reading Levinas

Edited by Robert Bernasconi and Simon Critchley

These essays provoke new responses to the work of the eminent French philosopher Emmanuel Levinas through an analysis of how the problematics of reading, deconstruction, feminism, and psychotherapy complicate and deepen Levinas's account of ethical responsibility. Published in English for the first time are two key texts in the Levinas-Derrida debate "Wholly Otherwise" by Levinas and "At this very moment in this work here I am" by Derrida.

Studies in Continental Thought

cloth $37 50

paper $14.95

The Jewish Bible after the Holocaust

A Re-reading

By Emil L. Fackenheim

". . . a carefully constructed, deep reading of the Bible for a post-Holocaust world; Fackenheim . . uses personal recollection, alternative scholarly interpretations, and a rich encounter with the Biblical text to say something important about Judaism today. . . The result is powerful and frequently moving "—Michael L Morgan

cloth $27.50

The Patient's Ordeal

By William F. May

In a series of moving studies, William F. May explores the ways patients and their families face their dilemmas and cope with their crises. A book of great humanity about people in difficult circumstances that speaks to us all.

Medical Ethics

cloth $24.95

The Patient-Physician Relation

The Patient as Partner, Part 2

By Robert M. Veatch

Robert M. Veatch proposes a postmodern medicine in which decisions about patient care will routinely involve both doctor and patient

Medical Ethics

cloth $27.50

Now in paperback!

Role Playing and Identity

The Limits of Theatre as Metaphor

By Bruce Wilshire

". . . Wilshire makes unique contributions to understanding major aspects of the human condition in its necessary search for selfhood."

—*Process Studies*

"It is one of the American classics."

—*Human Studies*

paper $14.95

cloth $29.95

Now in paperback!

Time, Narrative, and History

By David Carr

" a clear, judicious, and truthful account, provocative from beginning to end "

—*Journal of the British Society for Phenomenology*

" . . a superior work of philosophy that tells a unique and insightful story about narrative "

—*Quarterly Journal of Speech*

paper $10 95

cloth $22.50

INDIANA
UNIVERSITY
PRESS

Bloomington, IN 47406
Credit card orders call
1-800-842-6796

PHILOSOPHY *from* CORNELL

Socrates, Ironist and Moral Philosopher

By GREGORY VLASTOS. In this long-awaited work, Vlastos reclaims the remarkable originality of Socrates' thought. "A vivid and compelling account of Socrates' moral philosophy, and, more generally, of Socrates' moral outlook and his attitude toward religion and society." —Terence H. Irwin, Cornell University. *Townsend Lectures Series. Cornell Studies in Classical Philology*. 57.50 cloth, $16.95 paper

Identity\Difference

Democratic Negotiations of Political Paradox

By WILLIAM E. CONNOLLY. Connolly here explores some of the epistemic, moral, political, and social questions that the uneasy relationship between personal identity and difference poses for democratic politics. $27.50

Nietzsche's Voice

By HENRY STATEN. "An excellent piece of work offering a wealth of new insights. The author makes sense of more of the significant internal contradictions in the Nietzschean text than any previous commentator has done."—Graham Parkes, University of Hawaii. $29.95

Definitions of Art

By STEPHEN DAVIES. Davies describes and analyzes definitions of art in Anglo-American philosophy during the last thirty years, and in the process, suggests how we should reorient our thinking. $31.50 cloth, $12.95 paper

A Peculiar Fate

Metaphysics and World-History in Kant

By PETER D. FENVES. Fenves argues that, far from being a mere illustration of his metaphysical principles, Kant's attempt to present history in its entirety played a vital role in the transformation of his concept of philosophy. *A Peculiar Fate* demonstrates for the first time how Kant's concern with history motivates and gives shape to his "discovery" that a systematic philosophical inquiry must rest on human freedom. $36.50

Postmodernism and Its Critics

By JOHN MCGOWAN. Providing a comprehensive overview of the philosophical context of postmodernism, McGowan considers the kinds of freedom and oppositional politics possible under postmodern conditions. He focuses on the influence of Kant, Hegel, Marx, and Nietzsche and shows how the model of the artist as opponent of commercial culture has evolved from Romanticism to the present. $39.95 cloth, $12.95 paper

CORNELL University Press

124 Roberts Place, Ithaca, NY 14850

Cambridge University Press

Rationality in Action: Contemporary Approaches
Paul K. Moser, Editor

Twenty-one previously published selections concerned with rational action are listed under three main categories: individual decision theory; game theory and group decision-making; reasons, desires and intentionality.
Contributors: Paul K. Moser, R. Duncan Luce, Howard Raiffa, Leonard J. Savage, Daniel Ellsberg, Maurice Allais, Daniel Kahneman, Amos Tversky, Herbert Simon, Robert Nozick, David Lewis, John C Harsanyi, Robert Axelrod, David Gauthier, Kenneth J. Arrow, Amartya Sen, Brian Barry, Bernard Williams, Richard B Brandt, Robert Audi, Donald Davidson and Michael Slote
38572-5 Hardcover $59.50
38598-9 Paper $19.95

Human Nature and Historical Knowledge
Hume, Hegel and Vico
Leon Pompa

This book presents a study of the nature and conditions of historical knowledge, conducted through a study of the relevant theories of Hume, Hegel and Vico.
38137-1 Hardcover $44 50

Embryo Experimentation
Peter Singer, et al., Editors

Is embryo experimentation ethically acceptable and how should a democratic society deal with so controversial an issue, where conflicting views are based on differing religious and philosophical positions? This book is a current compendium of ideas and arguments on the subject.
Contributors: Stephen Buckle, Max Charlesworth, Karen Dawson, Beth Gaze, Richard Hare, Pascal Kasimba, Helga Kuhse, Peter Singer, Alan Trounson and Mary Anne Warren
38359-5 Hardcover $39 50

Hegel's Ethical Thought
Allen W. Wood

"This is a excellent treatment of a subject which has never been adequately dealt with in English…His book will be the most important study of Hegelian ethics since Bradley's Ethical Studies."
—*H.S. Harris, York University*
37432-4 Hardcover $49 50
37782-X Paper $16.95

Value and Justification: The Foundations of Liberal Theory
Gerald F. Gaus

"Gaus' work constitutes a highly original attempt to provide liberal theory with psychological and moral foundations."
—*Philip Pettit,*
The Australian National University
Cambridge Studies in Philosophy
37625-8 Hardcover $59.50
39733-2 Paper $19.95

Meaning and Speech Acts
Volume I: Principles of Language Use
Volume II: Formal Semantics of Success and Satisfaction
Daniel Vanderveken

In this book the author develops the logic of speech acts and the logic of propositions to construct a general semantic theory of natural languages. Volume I explains the general principles that connect meaning, reason, thought and speech acts in the semantic structure of language. Volume II uses the resources of philosophical and mathematical logics to develop a formalization of the laws of the semantic theory advanced in Volume I.
Volume I
37416-4 Hardcover $49.50
Volume II
38216-5 Hardcover $49 50

New in paperback…
Ecology, Community and Lifestyle
Outline of An Ecosophy
Arne Naess
David Rothenberg, Editor

Thoroughly updated by the author, this revised and expanded translation of Naess' *Okologi, Samfunn og Livsstil* conveys his thoughts on the relevance of philosophy to the problems of environmental degradation.
34406-9 Hardcover $44.50
34873-0 Paper $17.95

Cambridge University Press

New in paperback...

The Cambridge History of Renaissance Philosophy

C.B. Schmitt, Quentin Skinner, Jill Kraye and Eckhard Kessler, Editors

"In selecting Charles Schmitt as general editor...Cambridge University Press guaranteed originality and innovation... On the basis of the soundness of the general editor's overall plan and the strength of his main contributors the *Cambridge History of Renaissance Philosophy* is an outstanding success."

—*Charles Webster, Times Higher Education Supplement*

25104-4 Hardcover $82.50
39748-0 Paper $29.95

New in paperback...

A History of Jewish Philosophy in the Middle Ages

Colette Sirat

"Mme. Sirat proves to be a trustworthy and authoritative guide. She relies to a large extent, quite rightly, on lengthy quotations from the philosophers she discusses, and she manages to place them neatly and accurately in wider sociological and intellectual contexts."

—*Jewish Chronicle*

Copublished with the Maison des Sciences de l'Homme
26087-6 Hardcover $67.50
39727-8 Paper $22.95

The Call to Personhood

A Christian Theory of the Individual in Social Relationships

Alistair I. McFadyen

"What is a person?" Although the answer is given in largely theoretical terms, the primary concern of this book is to identify the personal, social and political practices required to live as a human in community with others.

38471-0 Hardcover $49.50
40929-2 Paper $14.95

Fallen Freedom

Kant on Radical Evil and Moral Regeneration

Gordon E. Michalson

This work offers a clear exposition of evil and moral regeneration as they appear in Kant's late work *Religion Within the Limits of Reason Alone.*

38397-8 Hardcover $34.50

Metaphor and Film

Trevor Whittock

Trevor Whittock demonstrates that feature films are permeated by metaphors that were consciously introduced by directors. An examination of cinematic metaphor forces us to reconsider the nature of metaphor itself, and the ways by which such visual imagery can be recognized and understood, as well as interpreted.

Cambridge Studies in Film
38211-4 Hardcover $39.50

The Rhetoric of Berkeley's Philosophy

Peter Walmsley

This study assesses Berkeley as a writer, offering rhetorical and literary analyses of his four major philosophical texts.

Cambridge Studies in Eighteenth-Century English Literature 6
37413-8 Hardcover $42.50

Demystifying Mentalities

G.E.R. Lloyd

The notion of distant mentalities has been used extensively by historians to describe and explain cultural diversity. Professor Lloyd rejects this psychologizing talk of mentalities and proposes an alternative approach, which takes as its starting point the social contexts of communication.

Themes in the Social Sciences
36661-5 Hardcover $49.50
36680-1 Paper $14.95

Available in bookstores or write:

CAMBRIDGE UNIVERSITY PRESS
40 West 20th Street, New York, NY 10011. Call toll-free 800-872-7423
MasterCard/VISA accepted. Prices subject to change.

Oxford

Equality and Partiality

THOMAS NAGEL, *New York University*

Derived from Thomas Nagel's Locke Lectures, this book proposes a nonutopian account of political legitimacy, based on the need to accommodate both personal and impersonal motives in any credible moral theory, and therefore in any political theory with a moral foundation.

May 1991 224 pp. $16.95

The Problem of Evil

Edited by MARILYN McCORD ADAMS and ROBERT MERRIHEW ADAMS, *both of University of California, Los Angeles*

Including an introductory essay and a selected bibliography, this comprehensive and completely up-to-date collection is an invaluable guide to current scholarship in this highly debated area of the philosophy of religion.

1991 240 pp.
paper $15.95 cloth $59.00

Free Will and the Christian Faith

W.S. ANGLIN, *McGill University*

In this book, Anglin demonstrates that free will and traditional Christianity are compatible. He examines, and solves, puzzles about the relationships between free will and omnipotence, omniscience, and God's goodness, using the idea of free will to answer the question of why God allows evil, and presenting arguments that link free will to eternal life and to the nature of revelation.

1991 232 pp. $45.00

Cartesian Method and the Problem of Reduction

EMILY R. GROSHOLZ, *Pennsylvania State University*

This book examines in detail the negative and positive impact of Descartes's method on his scientific and philosophical enterprises, exemplified by the *Geometry*, the *Principles*, the *Treatise of Man*, and the *Meditations*.

1991 176 pp. $39.95

Adam Smith and His Legacy for Modern Capitalism

PATRICIA H. WERHANE, *Loyola University of Chicago*

Werhane argues that a careful reading of Smith's major works show that it is justice, not self-interest or benevolence, that is the most basic virtue to Smith, and that a system of natural jurisprudence is necessary for a viable as well as an ideal political economy.

June 1991 224 pp. $29.95

New in paperback!

Philosophy, Psychiatry and Neuroscience—Three Approaches to the Mind

A Synthetic Analysis of the Varieties of Human Experience

EDWARD M. HUNDERT, *McLean Hospital and Harvard Medical School*

"Ambitious, rich, detailed, and groundbreaking."—*Review of Metaphysics* "A bold, challenging, exciting book...that integrates philosophy, psychology, and neuroscience to address the debate between rationalists and empiricists about the content and acquisition of categorical concepts."—*Philosophical and Phenomenological Research*

Paper 1991 360 pp.
paper $19.95 cloth $49.95

Prices are subject to change and apply only in the U.S.
To order, send check or money order to: Humanities Marketing, Dept. MG
To order by phone using major credit cards please call
(212) 679-7300, ext. 7106

Oxford University Press
200 Madison Avenue • New York, NY 10016

JUL 1991

THE JOURNAL OF PHILOSOPHY

Extra Copy

VOLUME LXXXVIII, NUMBER 5
MAY 1991

page
225 *The What and the How* Joseph Almog
245 *'P, therefore, P' without Circularity* Roy A. Sorensen

BOOK REVIEWS
267 G.A. Cohen: *History, Labour and Freedom: Themes from Marx* Andrew Levine
275 Arthur Fine: *The Shaky Game: Einstein, Realism, and the Quantum Theory* R. I. G. Hughes

280 NEW BOOKS: TRANSLATIONS

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1991 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NO. 5, MAY 1991

THE WHAT AND THE HOW*

The truth about Nastassja is this: she is a human being; she is also an actress, herself born to a famous actor. Discovered at a tender age by Roman Polanski, she was to be embroiled in much controversy that soon turned to fame. The French film critics, otherwise notorious for their iciness, describe her, breathlessly, as the long awaited heir, in both style and substance, to the legendary Ingrid Bergman. "Divine and sublime," wrote one of them reviewing her last appearance in the Cannes festival; "Sublime and divine," echoed another. In a more philosophical mood, and lest it be Griceanly inferred that the traits following the above "also" are meant to be modally contingent, I note that Nastassja is a twentieth-century figure, is not yet thirty years old, is Caucasian, and is of German descent; all are modally essential to her, if current essentialist doctrines are to be trusted.

---

* Work on the idea of logical truth which I had undertaken in the mid-1980s was goaded by *Tractatus* 5.552: "Logic is prior to the How, not to the What." Alas, I never really got, in that work, to reflecting on the distinction alluded to by this line, let alone to why logic would be prior to the one but not the other.

I have come upon the same distinction through a different lens, work on Descartes's discerning of mind and body. While attempting to reconstruct an argument for their discernibility that is neither from how we know the interested parties nor from alleged modal possibilities open to them, the fundamentality of the purported conceptual distinction grew on me. Sometime in 1987, I decided to compose an elementary exposition of this distinction, as I saw it. There resulted the present work.

I owe many thanks to Roger Albritton, Keith Donnellan, Y. Gutgeld, W. D. Hart, A. Hsu, and P. Leonardi. Especially helpful were David Braun and David Kaplan who, through close readings of early drafts, provided a constructive proof of how the intended distinction may be mistaken for others.

I refer to Bertrand Russell's *Introduction to Mathematical Philosophy* (London: Allen Unwin, 1919) as IMP.

0022-362X/91/8805/225–44

Truths about Nastassja they all are. But there is a difference in what *makes* them, in the actual world, into such truths: that Nastassja is human is true solely in virtue of *what* she is; the rest are made true not by what but merely by *how* she is. This notion, truth in actuality solely in virtue of what the subject is, involves, at least at first blush, neither mention of how we know the truth nor reference to its stability in counterfactual worlds. Later blushes are to confirm this: we have on our hands an autonomous notion of distinguished truth. I call truths solely in virtue of what their subject is *primal truths*. The elucidation of this notion is what this paper is about.

I THE INTUITIVE BASIS OF PRIMAL TRUTH

I.1 The coherent-subtraction test. Try to subtract from the above profile the fact that Nastassja is actually human, i.e., human in the actual world. For that matter, try to subtract from Nanga Prabat's profile its being a mountain, from Venus's its being a planet, from the Mona Lisa's its being a painting, from the Notre Dame's its being a cathedral. In the same vein, set yourself to tinker with the profiles of slightly more "abstract" subjects, e.g., the species whale and snail, the substances gold and water. Deprive the first of its being mammalian, the second of its being molluscan, the third of its being an element, and the fourth of its being a compound. All the attempted subtractions seem incoherent. Why?

Although Nastassja is my favorite subject, allow me, in view of the delicacy of this particular case, to begin with a more robust paradigm. Let it be the Nanga Prabat. In subtracting a trait from the profile of such a particular subject, e.g., 'being snow capped', we must strike a delicate balance between (i) merrily stripping away the subject of its predicates and (ii) preserving its integrity, viz., the sense that it is still this subject from whose profile we subtract. We can subtract with impunity snow-cappedness from Nanga Prabat's profile because, while we so falsify the record as to *how* this subject is, our sense of *what* is thus misdescribed remains intact. Intact it remains because our presuppositions, those which cement the very sense that it is this particular item with which we tinker, have been vigilantly kept out of the subtractive melee.

Alas, sometimes greed gets the better of us and our subtraction experiments violate the very presuppositions that make them possible. Attend again to the Nanga Prabat. Whatever liberties we are about to allow ourselves, if that worldly peak itself is to be the subject of our subtraction experiment, we must not tinker with its being a mountain. But then what of one who claims that 'being a mountain' is subtractible, too, offering as his "proof" the coherence of a supposition, a "short story" if you will, in which N.P. itself bears proudly the predicate 'not a mountain'?

The predicating hand destroys what the one handling the subject took so much care to protect. It is shamelessly supposed of our subject, a mountain if nothing else, that it is not one. Inevitably, the suspicion grows that our subject is not Nanga Prabat but merely an "arbitrary" entity, with no identity, a close kin to the philosopher's mysterious "bare particular." And so, with our most fundamental presupposition violated, the aboutness of the subtraction experiment slips away. Gone is the sense that it is still about the real Nanga Prabat, not an invented philosophical fiction, that we suppose.

Call the question—Is it coherent to subtract from x's actual profile the trait F?—*the coherent-subtraction test.* Four basic guidelines about how it should be read are in order. First, I should like it noted that it is only a test, viz., a method for identifying primal truths, rather than an analysis of the notion or an explanation of what *makes* a truth primal. In particular, the dependence on the act of subtraction and on what the subtractor(s) can or cannot mentally subtract should not be taken as constitutive of x's whatness. In agentless primordial times, 'mountain' already did and 'snow capped' already did not pertain to what the Nanga Prabat was (and is).[1]

Coherent subtraction is only a test. What is it a test *for*? The test seeks to discover whether a given trait F in x's profile pertains to its very whatness. Equally important, however, is to note what it is *not* a test for; the test leaves us free of a variety of (i) semantical, (ii) epistemic, and (iii) modal involvements. Thus, the second guideline is that the present metaphysical speculations about the nature of things, e.g., Nastassja, should not be confused with semantical hypotheses about their names, e.g., 'Nastassja'. In claiming that 'human' is a primal trait of N, and that the singular truth, 'that she is human', is primal, I have not claimed, e.g., like Peter Geach, that the English sentence 'N. is human' is true in virtue of meaning alone ("analytic"), true in virtue of the supposed "connotation" of the name. Nor did I claim that, if a user of the name 'Nastassja' failed to believe its referent was human or believed it was not human, he would fail to refer to Nastassja. I am here concerned with the nature of the object itself, the character of the bond between word and object notwithstanding.

The third guideline is that the test does not involve us in epistemological conjectures regarding what we may or may not know about the subject. I take it as given that Venus is in actuality both a planet and of interest to the Russians; also, that whales are both mammalian and of interest to marine biologists. I claim that in both cases we can coherently subtract the second trait but not the first. To stress the epistemic-free character of this claim, let it be granted (though, in the mantle of an epistemologist, I would not so grant) that (a) it is consistent with what we know that Venus is not a planet and whales are not mammalian; also, the even stronger (b) that it could turn out that Venus is not a planet and whales are not mammalian. To put the matter in even less hypothetical a form, consider a sufficiently ancient counterpart of the present author, also bent on discovering the what and the how of Venus and the whales. He is on record for suggesting that 'star' and 'fish' cannot be subtracted from their respective profiles. Led astray by his fellow astronomers and biologists, our ancestor metaphysician did turn out to be wrong (only his *meta*physics was, as it were, right). This epistemic fragility of his and my pronouncements, however, is quite irrelevant to the application of the subtraction test.

---

[1] The ramifications of the idea that the subtraction test is only an identification procedure for primal truth are discussed in detail in part IV.

The epistemologist's most prized target is the ultimately true profile of, e.g., Venus. He constantly operates with a merely temporary profile, worrying whether everything with it is quite what it seems. In contrast, the ancient metaphysician and I get hold of the epistemologist's target by stipulation, i.e., we begin with Venus's actually true profile as given. By successive subtractions of its merely descriptive traits, the how it is, we hope to unravel its predescriptive traits, those which guarantee the availability of a subject, the "what," for the eventual description. As such, the subtraction-receptive traits might not be epistemically inferior. They might, if, e.g., we focus on phenomenal traits of the right kind, be known to us with perfect certitude. Conversely, the subtraction-resistant traits are not, as already stressed in the case of Venus and the whales, epistemically superior. Subtraction resistance and receptiveness are determined by the nature of the subject investigated, not by the extent and hardness of the information available to the subtractor. We discover what Venus is by subtracting; what it is determines what we can subtract.

Finally, the fourth and last guideline: the test is meant to be taken in a modal-free way. I have just emphasized that we take the way Venus actually is as given and not worry about how it might turn out. This may encourage some readers to interpret the test as trying, given the way Venus is, to "lasso" the ways it might have been or, worse, to separate, among its necessary traits, the fundamental and the derivative. This is not the way in which the test is to be taken. The modal misinterpretation being the most prenicious, a few conceptual distinctions must be introduced at this stage.

I.2 Counteractual subtraction. Our test involves what I will call *counteractual subtractions,* not modal assertions of possibility. Counteractual subtractions and modal assertions differ, as we shall discover, in many ways. At this early stage, I merely observe that the two notions differ already at the level of extension. First, modally contingent propositions might be resistant to counteractual subtractions. This much was clear to Descartes (to whose insightful employment of the apparatus of subtraction experiments I shall return). He knew all too well he was not a perfect being who existed necessarily. And yet, he insisted, in the first person, that one cannot "subtract away" one's actual existence. I am not interested here in the epistemic dividends afforded by the first-person exercise; I am rather focused on the logical point implicit in Descartes's discovery: to subtract from the profile of any actual thing, other minds or other bodies, Nastassja or the Nanga Prabat, the trait of existence is to let the lively presence of the subject flatly contradict the supposed success of the subtraction. Some readers will nonetheless protest that we say nonchalantly, "Suppose that Nastassja didn't exist. . . ." Of course we do; however, what we so do is consider a merely possible history of the world, doing so all the while with Nastassja actually on hand. We consider the world as it was twenty-seven years ago and note that, for all but the grace of God, Klaus, that most notorious of womanizers, might have left off a month earlier with his next conquest, depriving us forever of the Nastassja we did come to enjoy. We can generate this much counterfactual history *of hers* only because *she* is actually on the platform.

The converse combination of coherent, counteractual subtractions that are modally necessary is equally available. This is so if only because for each trait F Nastassja actually has, N. is actually F is modally necessary. But for many such Fs, e.g., being an actress, we can suppose Nastassja actually F-less. "Actually" free cases make the same point. As modal assertions, the propositions that N. is a twentieth-century figure and that N. is human are equally necessary, if current modal metaphysics is to be our guide. In contrast, whereas counteractually subtracting the former is perfectly coherent, if not downright intriguing, dispensing with the latter leads to the irreversible loss of the subject of the whole exercise.

I.3 Existence conditions. The coherent-subtraction test is only a test. It provides a way of "lassoing" the extension of the notion of primal truth. The notion itself is not thereby analyzed. For myself, I harbor much skepticism when it comes to definitional analyses of fundamental (philosophical) notions. The aim of the remarks that follow is to supply an intuitive elucidation of the notion of primal truth. My basic idea is that, by reflecting on the coherent-subtraction test, we can distill such an elucidation.

To begin with, we need to distinguish, for any given subject, *conditions of existence* and *modes of existence*. The distinction is between traits the possession of which the subject's actual existence is conditional on, and traits the possession of which is conditional on that subject's actual existence.

More austerely, call a trait P a condition of existence for x iff it has to be actually true that, if x exists, x is P. P is a mode of existence of x iff P is a trait of x and is no condition of existence. A trait is a primal trait of x iff it is a condition of existence. A truth of the form x is P is primal iff P is a primal trait of x.

In line with the separation of counteractual subtractions and modal assertions of possibility, I note that the notion of a condition of existence is not to be confused with the modal notion of essence: conditions of existence are meant to resist counteractual-subtraction experiments, not refutations in arbitrary counterfactual worlds; they state, not what it takes for x to exist in an arbitrary possible world, but what is a *sine qua non* for x's existence in the actual world.[2]

My formulation relies on a seemingly mysterious 'has to be actually true' (though, again, arguments "by ear" do not convince me that in English it obviously means what we intend by '$\Box$'). Simple applications of the test of refutability by subtraction experiments versus refutability in an arbitrary counterfactual world should convince us (prior to much theoretical footwork that is to follow) that 'has to be actually true' differs already in extension from the 'has to' of modal necessity. As mentioned, 'N. is human', 'N. is a twentieth-century figure', and 'N. is actually an actress' are all modally necessary but only the first has to be actually true. On the other hand, there are propositions that are modally contingent but have actually to hold, if she is actually to be. 'N. exists' was already mentioned and 'If N. exists then' ('if N. is actually an actress then N. is an actress') is another such case: the proposition has to be actually true, as any other *F* we care to substitute for the trait of being an actress will verify. Nonetheless, the proposition is contingent as attested by a counterfactual world where Nastassja does not pursue a career in the movies.[3]

---

[2] There is, I believe, a traditional use in philosophy of the term 'essence' that is modal-free and is very much in the spirit of what I intend by 'condition of existence'. Perhaps even the ordinary language use of 'essential' can be so heard. In view of the dominance of the modal reading of 'essence' in current discussions, however, I shall so use it. Like remarks apply to the classical logical idea of 'necessary condition for . . .'. I hear it as a modal-free notion. But to avoid any strain *vis-à-vis* current use, I have adopted '*sine qua non* for . . .' where no alethic modality shows up.

I.4 Existence conditions and whatness. A proposition of the form '*x* is *P*' is a primal truth iff *P* articulates a condition of existence of *x*. But why, it may be asked, does the idea of condition of existence capture the primordial idea of what *x* is? In focusing on what, as opposed to how, Nastassja is, we proceed, as already suggested, by "analysis": beginning with the full profile of her traits, we go on to subtract the traits that describe her till we reach her core traits, those which guarantee the provision of a subject, the "what," for the eventual description. In contrast, in articulating existence conditions for N., we proceed by "synthesis": we pretend that N. is not around and wonder, now that we play the role of "the maker," what is absolutely a *sine qua non* for us to do if we are to synthesize her.

Of the two, the "synthetic" idea of existence condition runs deeper. It is because a trait is a condition of existence that it is predescriptive, not the other way round. It is written into the very idea of "description of a subject" that there be a subject to be described. Thus, if a trait is an existence condition, it does not belong with a description of how the given subject is but rather is part of the predescriptive stage setting: the provision of a subject, the what, whose acts and deeds are to be described. What is more, even when we engage in a deliberate attempt to specify N.'s "what" in existence-free terms (as in the above "analytic" procedure), allusion to her existence conditions is inevitable. We are to "peel off" her traits till we reach what may be termed "the critical point," the point where we do not dispense anymore with her being thus-and-such but her being at all. Her "what" is thus inextricably linked to what it takes for her to be.

---

[3] Another common misconstrual of 'has to be actually true': in oral presentations of the present material, 'has to be actually true' has, at times, been read as 'is eternally true'; this was "backed up" by some with the, allegedly "Aristotelian," idea that what I am after, viz., what Nastassja is, consists first and foremost of those features which persist with her through changes. Let me then be crystal clear about this "temporal interpretation." First at the level of the notions involved, read 'what N. is' not as 'what N. was, is, and will be', take 'what N. is' for what it is, viz., with a tensed 'is'. Also, nothing in my corresponding, logical 'has to' alludes to truth at (all) times. It is consistent with everything I have to say that one contests the very idea of having a property at a time (and not another) and of a proposition true at one time and not another; even if this much is not contested, it is still consistent, in a "Quinean" mold, to contest the very coherence of predicating individuals, in and of themselves, regardless of how they are specified, with tensed properties; finally, even if the intelligibility of tensed predication is accepted, it is still consistent to hold the view that Nastassja never changes, that all her features are temporally eternal. This may be because she is taken to exist for, say, 70 years, but without any changes or because her life is reduced to a very short period, in which no changes occur, the subsequent acts and deeds being really of a series of connected, but distinct, individuals. So much then for the independence of the notions. In addition, I should like it noted that, even if one were to accept a "standard" view of temporal notions (I do), there would still be marked differences at the level of the extensions of the two notions, viz , eternal truth is neither sufficient nor necessary for whatness. 'Was conceived on a rainy day', 'was not really planned', and 'was born in Germany' are all eternally true of N.; if nontensed predicates are wanted, 'has a birth mark on her back' and 'has a liver' will (probably) do; none of those pertains to what she is. Also, it is not inconceivable for me that, given radical changes in the environment, taking place in a gradual way, N. could turn from being human into being some other kind of higher sentient mammal. To rule out this possibility would take some argumentation going much beyond the question of what she is (now).

II. THE KIND OF QUESTIONS "WHAT QUESTIONS" ARE: ONTOLOGY

II.1 Classification versus individuation. The history of attempts to address "what questions" is long and distinguished; it may be part of the very *image de marque* of philosophy itself (is not many an immaculate student initiated to a vivacious Socrates hobnobbing with marketplace cobblers and pondering with them the question, 'what is a shoe?'?). Reflection on the labors of such distinguished what-seekers reveals that many of them heard in the question, 'what is x?' a call not merely for existence conditions but for full identity conditions, viz., answers of the form, 'why, but it is the unique *F*', where '*F*' specifies the "real definition" of *x*.

By this, philosophers have not merely aimed at (modally) necessary and sufficient conditions for being Nastassja: witness the summary dismissal of 'the actually favorite actress of Roman Polanski' or 'the sole member of {Nastassja}'. Similarly, Richard Dedekind, G. Peano, Gottlob Frege, and Bertrand Russell, all intent on articulating what, e.g., the number 4 is, would not allow, even if they were to swallow the modern infatuation with modal essentialism, the answer, 'number that solves the map-coloring hypothesis', or even, 'number of dimensions of a Minkowski space', modally necessary and sufficient as these traits might be. What such what-seekers look for is a specification that articulates what *x*'s identity "consists in." For example, if some philosophers got their way, 'such and such continuous succession of human stages' would do in N.'s case; 'concept any first level concept falls under when four members fall under the latter', 'set of all four-membered sets', and 'successor of 3' did, *grosso modo,* the trick for the aforementioned number analysts. Responding, as they do, to 'what is *x*?' queries with individuating specifications, I shall call such what-seekers *individuation theorists.*[4]

---

[4] I have in mind a whole variety of individuation theorists. All seem to me moved, *au fond,* by "reductionist" tendencies: truths about the original subject

From the beginning of this paper, I have adopted the mantle of a classification theorist. I have presupposed, rather than argued for, my view that in articulating the whatness of many worldly subjects, as natural as Nastassja and the Nanga Prabat or as artifactual as the Mona Lisa and the Notre Dame cathedral, "defining," better called "individuating," features need not be included. This much does not deny that some trait '*F*' may uniquely apply to them, perhaps even in all possible worlds. I rather predict that such a uniquely specifying trait will merely describe how they are, perhaps how they are in all worlds, but will not divulge *what* they are. In a less presuppositional mode, I note that in response to 'what is Nastassja (Nanga Prabat, Mona Lisa)?' we do not despair and even appreciate 'a human being', 'a mountain', and 'a painting'.

In previous work, I have been inclined to challenge the individuation-oriented approach. Not here. Doubtless, the individuation theorist and I supply different answers to 'what is x?' queries; we agree, however, on the kind of question it is and, equally significant, the kind of question it is not. I have come to believe that the prize for abstracting from such "local disagreements" is double: first, a sharper, dare I say clear and distinct, perception of the kind of question 'what is *x*?' is; secondly, a better understanding of where the two pictures agree, where not, and, most important, which of these matters more.

---

can (and should) be "reduced" to truths about more "basic" items, viz., those in terms of which the individuation criteria are stated. Talk of "reduction" comes in various shades and half tones. Some will want the original subject, e.g., N , the number 4, the species whale, literally identical to certain equivalence classes (of her stages, of four-membered sets, and of animals descending from a common origin); others will avoid identity talk and seek "contextual definitions" that eliminate the merely apparent subject, in reality a "logical fiction," in favor of purely quantificational claims (Russell, at one point, thought almost everything under the sun, even the sun itself, was such a logical fiction); yet other theorists would favor the "disappearance theory" of Nastassja et alia: there is no such thing as N., we are in the grip of a "perceptual error" (perhaps Hume's theory of the self can be so read). Milder positions would preserve Nastassja's integrity. It is simply that truths about her may be taken to "supervene" on truths involving her stages (the terminology of "supervene" allows itself many different readings); the equivalence class may be taken to "represent" Nastassja; and so on.

Yet another dimension of variation among individuation theorists is the one involving extensional and intensional individuation theorists. The extensionalist correlates one subject (e.g., Nastassja) with another, extensionalized subject (e.g., the equivalence class) The intensionalist correlates a subject (e.g., Nastassja) with a predicate, viz., a condition specifying uniquely what N is (e.g., 'continuous succession *S* of human stages') In the text, I shall not engage in subtleties arising from the existence of various clans of individuation theorists. My remarks are framed to cover them all.

II.2 Ontological questions: the autonomy of primal truth. What kind of trait do we ascribe to a truth when we pronounce it "true solely in virtue of what its subject is"? The trait so ascribed and the question thus addressed, whether in the manner of the individuation or the classification theorist, are, I believe, rather distinct. I shall call such questions *ontological*. In what follows, their autonomy and, in particular, independence from the more familiar epistemic and modal queries, will be set forth.

Paradigmatic to epistemic questions about Nastassja is the search for propositions that have to be true because of the *way* we know that they *are* true. So read, 'it has to be true that N. is self-identical' (compare: 'that I think'), is true because of the way in which we know that the embedded claim is true. Paradigmatic to modal questions about Nastassja is the search for propositions that are true no matter which way the world might have been. So read, 'it has to be true that N. is self-identical' (compare: 'that I am the offspring of gametes *J*') is true because the embedded claim could not have been false.

Ontological investigations concern neither our knowledge of the actual world nor the truths of the nonactual ones. They concern the actual *world*, the things themselves, and they concern the *actual* world, not the merely possible others. Of course, the ontologist does not care to investigate whether N. brushes her teeth or whether the Nanga Prabat has a helicopter at its very top. The *state* of things is not the ontologist's business; the things are. He strives to articulate the make up of this world, what there is and what it is that is.

So understood, the what-seeker, individuation or classification oriented, is in search of propositions that have to be actually true, if their subject is actually to exist, if there is to be this "what." Read this way, 'it has to be true that Nastassja is self-identical' (compare: 'that, if N. actually brushes her teeth, N. brushes her teeth') is true because, if she is to come into actual being, her being self-identical is a *sine qua non*.

The autonomy of primal truth and its irreducibility to notions of distinguished truth of the classical variety calls for some additional observations, equally precious to the classification and individuation theorists. Can truth solely in virtue of what N. is be a disguised form of truth solely in virtue of our internal "concept" of N.? It seems not. Apriority, however we understand this intractable notion, is not coextensive with primal truth. We have only to recall the primality, agreed upon by the classification and individuation theorists, of, e.g., the a posteriori truths that Nastassja is human, Venus is a planet, whales are mammals, and gold is an element. Follow an individuation theorist like W. O. Quine, and truths like 'Elmer is such and such a continuous succession of tree stages' are classed definitive, true solely in virtue of what Elmer is. But as Quine himself is the first to point out, the establishment of such truths involves investigations of processes of "chemical deformation," of what is involved in being a tree, etc., all empirical matters par excellence.

A more prevalent suspicion is that, in classifying a truth as primal, we ascribe to it a modal distinction of being necessary and of so being because it specifies the subject's modal essence. The suspicion is unjustified. I have already mentioned propositions like 'N. is a twentieth-century figure', 'N. had egg E at her origin', 'N. is actually an actress'; 'The species whale originated before 1066AD', and 'gold has specific gravity larger than 1' can be added to the list. Accept current essentialism and necessary they all seem; articulators of what their subjects are they are not.[5]

Is involvement with modality at least necessary for the primalist? One has only to recall the Frege-Russell search for what the numbers are coupled with their disdain for the modal to start wondering. Let yourself be a *de re* modal skeptic: predicating of Nastassja necessary humanity, in and of herself, is meaningless. In the language of "possible worlds," you hold, whether for Quine's, Alonzo Church's, or David Lewis's reasons, that the idea of nonconventional identity across worlds is bogus. This much does not bar you from accepting that, in the actual world, 'human', and even 'continuous succession *S* of human stages' if you sympathize with the individuation theorist, pertain to what N. is, 'actress' merely to how she is. In fact, you may even radicalize your modal skepticism so much as to reject what Quine calls the (mere) "second grade involvement with modality," viz., for you necessity cannot even apply to closed sentences ("general propositions") of the object language; better confine it to the metalanguage (and read it as 'is analytic'). So be it. You can still accept that 'human (continuous succession *S* of human stages)', not 'actress', gets at what Nastassja is.[6]

---

[5] Another way of making the same point: suppose that, like some fatalists and other modal determinists, you believe that every trait N. happens to have is hers of necessity, that things could not have been otherwise. Even in this, "deterministic" setting there are necessities and there are necessities: that N. is an actress is a necessity pertaining to how she is, that she is human to what she is.

[6] Some readers may still inquire: Will not all primal truths be necessary for those who do acknowledge the meaningfulness of *de re* modal predications? If so, would not this show that the notion of primal truth has a modal underside after all? We should separate here two questions, a separation exemplified by the case of the analogous thesis, "All known propositions are true " First, there is a ques-

II.3 Individuation versus classification: a larger perspective. A modal metaphysician who assents to the necessity of Queen Elizabeth's being a caucasian but questions the necessity of her origin or assents to the necessity of this artifact's being a table while questioning the necessity of its being wooden, need not, most probably does not, question Saul Kripke's *notion* of necessity. He is simply more modest about its extension. Similarly here: the individuation and classification theorists share their notion of primal truth. To wit: a proposition 'x is P' is a primal truth iff it has to be actually true that if x exists then x is P. What they differ about is the status of some such Ps.

I shall not try to convince the reader that my more modest extension is the correct one. What is more, which extension one ultimately settles for may be secondary in importance to a proper understanding of the notion involved. The individuation and classification theorists share the larger picture: (i) they agree on the kind of question 'what is x?' is and on its autonomy *vis-à-vis* modal and epistemic matters, (ii) they agree on the pretheoretic intuition that propositions like, e.g., 'N. is human', 'Cayster is a river', and 'Whales are mammals' are true solely in virtue of what, not how, their subjects are. Pretheoretic agreement on such paradigm cases and on the basic notion involved—a trait x has actually to have, if x is actually to be—may be of more significance than differences over controversial cases motivated, in the first place, by theoretical desiderata.[7]

---

tion about the extensions of the relevant notions, viz., do all known propositions happen to be true? The answer is 'yes'. Quite a different question is about the two notions: Can sheer conceptual analysis of the subject notion, "known proposition," reveal its subsumption under the predicate notion, 'true proposition'? The answer is again 'yes'. In our case, things are rather different. As noted in connection to *de re* modal skeptics, conceptual analysis of the notion "primal truth" will not certify its subsumption under "necessary truth." If anything, it is by reflecting on our notion of necessity that one may argue that, when projecting necessities about x, we constrain them by truths about what x is. Far from exposing a modal underside in the notion of primal truth, the notion of necessity would be found to have a primal underside. Now these observations are all on the assumption that, at the level of extension, all primal truths are indeed necessary. Are they? According to the next section, logical truths are a special case of primal truths. In my "Logic and the World," *Journal of Philosophical Logic* (1989), contingent logical truths have been presented; hence there would be contingent primal truths. Are there then nonlogical contingent primal truths? I cannot address this matter here, involving as it does a consideration of propositions more complex than subject-predicate propositions and hence, a violation of the limits I set myself in this elementary exposition. I hope to return to this in future work on the very idea of necessity

[7] (To the technically minded reader.) One additional way in which the classification and individuation theorists are brought here under a single roof is this: the

III. THE PRIMAL AND THE STRICTLY LOGICAL

III.1 Basic versus derivative primal truths. Much mental agility is called for tackling ‘what is x?’ questions. We do not want Nastassja and (the planet) Venus as same-whats just because, down the line, both are extended; nor do we want (the dog) Checkers and (the dolphin) Flipper as what-mates even though both are mammalian. On the other hand, most of us would not want to see Nastassja and Winnie (Mandela) segregated just because Nastassja is caucasian but Winnie is not; nor will we allow invidious distinctions in respect to whatness between Nastassja and St. Thomas just because she lives in the twentieth century and he lived a few hundred years earlier or because he was a "he" and she is a "she." We want Venus basically classified as a planet, Checkers as a dog, and St. Thomas, Nastassja, and Winnie as the same-whats: human beings.

A natural distinction between the *basic* and the *derivative* senses of ‘what is x?’ imposes itself. In its intended, basic sense, we aspire with ‘what is x?’ to the most minimal answer that is not too minimal. In contrast, when we respond with ‘a mammal’ for the last four mentioned items or with ‘an extended thing’ for all of them, it is merely a derivative sense of the question that is addressed. Yet again, the idea of such a distinction is acceptable to the individuation-oriented philosopher; he simply disagrees, in particular cases, on whether my candidate traits are basic enough.

Just as we distinguish basic and derivative senses of the question, I suggest to distinguish basic and derivative primal traits and truths. ‘Nastassja is human’ and ‘Venus is a planet’ are basic; there follows a

---

identity conditions of the latter are seen as a special case of existence conditions, viz., existence conditions that are uniquely specifying. This approach contrasts with the more standard view on which, if anything, existence conditions for x are seen as stated in terms of x's identity conditions, viz., it is required that x's identity conditions be "instantiated." On this last picture, the actual Nastassja (say, specified by ‘offspring of gametes N’) and her merely possible twin sister (‘offspring of gametes M’), have been equally provided with identity conditions; their difference lies in that the former's, but not the latter's, identity conditions have been instantiated. This contrasts with the picture I have presented which insists that there is all the difference in the world between what actually is and what actually is not. For those that are, there are existence conditions and, if the individuation theorist is correct, in some cases such conditions might be uniquely specifying. On the other hand, that which is not is not. There is no "it" for which conditions, individuative or classifactory, can be stated On my picture, "existence precedes identity," i.e., either x exists or not. If x exists, the truth that it does, "x exists," is basic and identity facts involving x presuppose it (to have x and y enter the identity relation, they must actually be). Thus, for x to satisfy ‘to be identical to y’, x has (i) to be and (ii) to be identical to y. On the other hand, if x does not actually exist, there are no truths involving "it", in particular no truths about what "it" is identical with nor truths about what "it" is chain of truths, all primal, all stating conditions that have actually to obtain if their subjects are actually to exist, e.g., 'humans are mammals', 'mammals are animate', . . . and . . . '*Xs* are extended'; or, 'planets are heavenly bodies', . . . and . . . '*Ys* are extended'. Consequently, 'Nastassja is extended' and 'Venus is extended' are primal truths but derivative at that. At the limit of such "derivations," we reach the strictly logical truths about Nastassja and Venus, e.g., each is, is self-identical, is red or not, is actually red if red, etc.

This last point relates to my introductory note and its record of an original hope to forge a connection between the primal and the "strictly" logical. On the traditional approach to the matter, such a project is discouraged from the very beginning. For we traditionally begin by investigating the strictly logical truths, attending to purely formal traits of Nastassja, traits she shares with all beings as such. Truths like 'Nastassja is human' are thus sentenced to remain out in the cold. On the present picture, we begin, classificationists and individuationists alike, with 'what basically is N.?'; so we also proceed with any other subject, Venus and Sirius, whales and snails. We then go on to more general traits shared by some of these subjects ('are animate'), to yet more general traits shared by all ('are extended'), eventually ending up with the strictly formal traits ('are self-identical'). All these propositions now naturally belong in the same kind, viz., predications that have to be actually true, if their subjects are to actually exist. The primal truths at the limit, the "strictly" logical truths, are seen to be "as concerned with the real world as truly as the truths of zoology, though with its more abstract and general features" (IMP, 169).

A pleasing, "technical" codification of the foregoing suggests itself. On the standard view, it is the idea of logical truth that is taken as basic. A primal truth is defined as the logical truth of a conditional 'If A then ϕ', with A some fundamental assumptions we hold fixed. On the present picture, the idea of logical truth emerges as the derivative notion, viz., ϕ is a primal truth closed under certain universal permutations of its individuals and attributes. This reversal of cart and horse carries some irony with it; and irony aside, it may hold the key for decoding the, at least a, thought behind "Logic is prior to how not to the what." First though the irony.

Logic, I here mean not so much what the mathematical logician practices but what the philosophical logician likes to theorize about, is often heralded as free of any ***existential involvement***. By this last intended are both the object-language thesis that singular existential truths recording what there is, e.g., our 'Nastassja exists', are the epitome of merely material, hence nonlogical, facts; also, the metalanguage thesis that logical truths have no business committing us to the existence of anything. Even Russell, barely a few pages (glasses?) after his bold worldly *leitmotif,* recants. First, we are told that, though the primitive propositions of *Principia Mathematica* allow the inference that at least one individual exists, he now (1918) "regards it as a defect in logical purity." Harsher even, "there does not seem any logical necessity why in fact there should be any world at all" (IMP, 203). This is the gentle news, what with Russell at least trying to fly the worldly kite, then getting cold feet. More middle-of-the-road writers would simply dismiss as a charming joke the idea that logic is as concerned with the real world as zoology.

Charming or not, a joke it is not, at least not by the present lights. To the philosophical logician, existence, used or mentioned, may be the ghost in the logic ("inference") machine. Quite the contrary, says I: what is and what it takes for it to be are the very *terrain de chasse* of logic. It hunts the most fundamental, abstract and general, features of real things and the world they make up; so fundamental, that without them gone is the sense that those things *are,* that the world they make *is.* This focus on what is and what it takes for it to be shows in what I think of as the logician's, not the zoologist's, trap, the aforementioned frame: "It has to be actually true that: if x exists, x is P." As before, there is the metalanguage side of the matter with the trap designed to catch what is a *sine qua non* for Nastassja, by which 'non' is meant no *being,* absolute nothingness, for the old girl. There is also the object-language aspect with '(actually) exists' primed as a most obvious value for 'P'.

I am now in a position to speculate on the connection between Nastassja's logical traits and what she is, a connection inspired by the mysterious yet seductive "logic is prior to the How, not to the What." As I see the connection, it is induced by two separate links—between what Nastassja is and existence conditions for her, a link discussed early in this work, and the just cited tie between her logical traits and, yet again, existence conditions for her.

Logic is seen to be prior to Nastassja's how not her what because, in studying what she is and her logical features, we are pursuing the same quest, viz., her *existence* conditions. It is simply that, in the logical case, the conditions are so abstract and general that they are universalizable. In both cases, we prowl prior to the how because we are focused on the primeval stage of the *provision* of subjects, not the eventual *description* of those subjects.

All in all, and lest there be any misunderstanding—my task is not to push the frontiers of logic so that they embrace Nastassja's humanity and other such treasures. 'Graceful-or-not' is logical and universal in extension, 'human' is not, and there is no question about this. The question is rather—should we canonize the extent of the extension of a predicate—is it universal or not—as our *quitte-ou-double* question? Elevating 'graceful-or-not' to a lordly position and consigning 'human' and 'featherless biped' to equal oblivion? Or should we rather start with the "meaning" of the predicate, I would prefer the property, involved, regarding all the while the extent of the predicate's extension as a corollary, flowing from how general and abstract a condition it is? I have been urging us this less trodden path.

IV. AN IDENTIFICATION PROCEDURE FOR PRIMAL TRUTH

IV.1 Cases. The notion of primal truth differs both qua notion and in extension from both apriority and necessity. Like primal truth, I believe that neither the notion of necessity nor that of apriority admit definitional analyses. What can and has been supplied for this pair of notions is a "computable criterion" for membership in their extension, viz., a method for "lassoing" the class of relevant truths. Such is the role of extensional surrogates, those which replace the "uncomputable" intensional original by a "computable" extensional surrogate, viz., " '*x* is *P*' is necessary iff '*x* is *P*' is true in all possible worlds" and " '*x* is *P*' is apriori iff '*x* is *P*' is true in any evidential situation (context)." Can we provide a similar, "extensional" and "computable," identification method for primal truth?

Consider the question: What do we *do* when faced with a claim that *x* has actually to be *P* if it is actually to exist, or, less theoretically, that what *x* is is *P*? I suggested that we engage in a counteractual subtraction of *P* from *x*'s profile and see whether it is stopped in its tracks. But what is a counteractual subtraction? *Where* do we separate *x* and *P* when we so subtract? After all, in the actual, real world, *x* is *P*. The answer is: we separate them in a *story*, in our local vernacular, in a *case*, one that has a *P*-less *x* as its leading character. Systematic failure to come up with such a countercase will lead us to suspect that there is truth in the what-claim. As I read our practices, they suggest that:

(P) '*x* is *P*' is primal iff all coherent cases about *x* make it *P*.

IV.2. Hard Cases. (P) is meant to settle what-conjectures by running through cases. Inevitably, some of the cases will turn out to be hard cases. To make no bones about the problem, consider what, in my eyes and to my knowledge, is probably the hardest case of all, viz., Descartes's question 'what am I?' and his subsequent story about himself, there by the fire, but his body, his very extendedness

I should say, shed away. A story he took to show, at the very least, that being extended is not part of what he is (and not merely, as the modalists instruct us, that it is not part of his "modal essence"). A coherent case? A hard case above everything else. Alas, some of us, perhaps with the ideal of a *Methodenreinheit* in mind, tend to leap from the existence of such hard cases to skepticism about the very method.

My apologia is that the purity, let alone power, of a philosophical method should not be evaluated with the false ideal of a universal-decision procedure in mind. Intrinsically, even if seemingly elementary, hard cases are just the natural breaking points of reliable methods. We do not dismiss our computing methods when a diophantine equation in 9 variables turns out to elude them. We do not dismiss the method because we begin to suspect that there may be no better one. This, in turn, forces us to reflect on the reasons that made this particular school-boy arithmetic problem recursively unsolvable; we then reflect some more on the nature of the method itself and come to see why it had to break down at this point: if it did not, something too good to be true would have been happening. We mature. We realize what the method can do and what no method can undo.

What is good for the number theorist is good for the philosopher: Descartes's case should come out a hard case, perhaps intrinsically unsolvable. Were the method to suggest that things here are crystal clear rather than as seen through a fog, I at least would be instantly suspicious of the method. Far from being an embarrassment, hard cases (of the Cartesian delicacy) provide the method with the seal of reliability. Instead of lamenting the existence of a few hard cases, we can bask in the verdicts delivered by most of them. Using his own favorite method, we can show that it is not part of what Descartes is to be a philosopher or/and a mathematician; nor are, and this is more intriguing a conclusion, the modally essential traits of being French, being Caucasian, living in the seventeenth century, and originating in egg D. In contrast, to be a thinking being and a human being are traits that are inseparable from him, so most of us feel, and thus should be counted as primal. The cases involving being in time, being extended, mammalian, etc., might be left unsettled. Not unsettled for me; but I should like it noted that we can give the method a clean bill of health even on the assumption that these cases are left undecided.[8]

---

[8] As hinted, here and below when the case of a painless c-fibers' firing is discussed, I believe that the present framework may enlighten us about the mind/

IV.3 Countermodels and counterexample cases. The method of cases is to the philosopher what the method of (counter) models is to the mathematician and the logician. All aim to perceive, clearly and distinctly, connections between propositions, in particular, whether a target proposition *S* is already "contained," or, in contrast, "separable from," a body of assumed propositions *B*. In investigating "containment" and "separability," the logician abstains from any modal involvement. Thus, the axiom of choice is ***modally*** dependent on the base axioms of set theory: in any world where the latter hold, it holds, too. Nonetheless, it is ***logically*** independent from those axioms, there being models verifying the base axioms but falsifying it.

Cases, the philosopher's models, display the same coupling of modal dependence with (onto)-logical independence: 'Descartes is a seventeenth-century figure' is modally dependent on 'Descartes is human'. Nonetheless, the former is ontologically independent of the latter, as a coherent story about that man living a century earlier or later testifies. All the cases fabricated in our search for what Descartes is may turn out to be genuine modal possibilities for him; nonetheless, it is not part of the point of spinning such cases to engage in the metaphysics of modality.

Cases can and do transcend "the modal horizon." To ground the point, it may be worthwhile to reflect on some classical cases. Suppose it turned out that the story about twin earth represented no genuine possibility for our universe. Perhaps because of the modal determinism of footnote 5, i.e., given that things could not have been otherwise and that there is no twin earth, there could not have been one; or as I prefer, simply because there is no point in the history of our universe where it could have gone on to produce an exact replica of our (already vastly improbable) earthly development. With twin earth representing no genuine possibility, would the point of Hilary Putnam's tale be gone by the board? Not as I read it. The point of the case is, to begin with, ontological, not modal, viz., for there to be water in actuality is one thing, any old liquid with waterish appearance quite another thing. This much is a

---

body distinction. I can even be more specific while remaining mysterious: the what-I-am versus how-I-am distinction affords a reconstruction of (one of) Descartes's arguments for the real distinction, a reconstruction that may well vindicate the intended distinction. At least so I argue in "Philosophical Fiction and Real Distinctions," the aforementioned work dedicated to such a reconstruction of Descartes. I hope to convince myself, sometime in the next few years, of the conclusions of my reconstruction.

premodal observation on what it takes for water actually to exist (its not sufficient that a waterish liquid shows up); this is not a modal claim about how our world, with the water it does actually have, might have been.

Like remarks apply to other classical cases; e.g., post-Kripkean critics have charged that a case he has presented, of Jones's c-fibers firing but no pain felt, may have no bite because a necessary connection, of course only empirically discoverable, may tie the firing of the fibers and the feeling of pain. In response, I say: let there be such a necessary connection; it does not rob Kripke of his ontological point, whether or not he himself intended to read the situation in this modal-free way. A coherent story can be concocted, it seems, about Jones's agitated fibers but no pain felt. This much has been conceded by physicalists who, in the old, pre-Kripke, days, agreed that there is nothing "conceptually incoherent" in a painless fibers' firing. I "translate" and concur: by 'conceptually incoherent', we mean neither "is a priori excluded" nor "is modally impossible." We mean rather "is barred by *what* the items are." Then indeed there is nothing in what c-fibers' firing is to bar a painless firing. In contrast, there is much to bar a firing that involves no atoms, is not located, consumes no energy, etc.

And so it goes: we are told not of one person but two, waking up two hundred years after the death of Guy Fawkes with memories (qualitatively) identical to Fawkes's; we are privy to Einstein's view of the universe as he is riding a light ray or the life and times of us, humans, in a two-dimensional space. With my guardedness over the span of the possible, I see in none of these stories a way the things themselves might have been. No matter. The point of these cases is not to speculate on the merely possible ways the world might have followed but to divulge secrets about the fabric of things here in actuality.[9]

V. A CONCLUDING NOTE

Cases provide us with "graphic evidence." They are the philosopher's brand of test tube, his way of probing the elusive chemistry of subjects like Nastassja. I remind the reader it is not the existence of a case, let alone its discovery, that *makes* a truth (non) primal. The critical question is whether P is an existence condition for x, i.e., whether it has to be actually true that, if x exists, x is P. At an even more primordial level, I would regard my own elucidation as somewhat derivative. At that ultimate level, we would find that we had come full circle back to the opening page of this work, where intuitions about what, as opposed to how, Nastassja is were alluded to. At this level, what makes an arbitrary proposition of the form 'Nastassja is *P*' true solely in virtue of what she is is what she is.

---

[9] Many, to judge by current metaphysical discussions, would allow that the cases represent no "metaphysical possibilities" but nonetheless are "logical possibili-

JOSEPH ALMOG

University of California/Los Angeles

---

ties." I confess that I am rather alienated by this popular mode of expression. In this paper, as elsewhere, I mean by 'necessary' "absolutely necessary," what, I think, Kripke means by his insistence on "necessity tout court." In particular, to recall the above example, the axiom of choice is necessary tout court (this is meant to exclude the Fraenkel-Mostowski countermodels, not just the later, pure sets, countermodels, as possibilities). To claim that the axiom of choice could have been, "logically" as it were, false, is, in my eyes, to confuse the metalanguage, nonmodal, claim, that there is a model of set theory, where it is false, with the object-level claim, concerning ways the world, even if abstract facets of it, might have been. Like remarks apply to the induction axiom and the existence of nonstandard models of arithmetic where it goes haywire. Formal independence proofs are simply no proofs of genuine possibility. Closer to home, recall that the axioms of propositional normal modal logics, as strong as S5, allow one-world models, models used in relative consistency proofs for such systems. But assume, as is only natural, that the metaphysics of our (fn. 5) modal determinist is false, that things could have been otherwise. Then by none other than the intended meaning of the S5 axioms, the one-world setting is not a genuine possibility. In my view then, much of the currently free talk of "logical" possibilities stems, ironically, from mistaking use and mention, the original sin in this land of confusion: consistency talk, which is (i) nonmodal talk about (ii) sentences (theories), is blended with (i') modal talk about how (ii') things, concrete or abstract, might have been. As I see it, models are models and possibilities are possibilities, the two simply belong in different strata. The idea that models are not modal and that, by simple independence results, there are many necessary (mathematical) truths for which countermodels are constructible is developed in detail in "Philosophical Fiction and Real Distinctions."

'*P*, THEREFORE, *P*' WITHOUT CIRCULARITY*

My main thesis is that a variety of arguments having the form '*P*, therefore, *P*' do not beg the question. A consequence is that syntactic theories of circularity are false and pragmatic ones are confirmed. A third theme of this paper is that some of these arguments owe their surprising persuasiveness to a neglected feature of arguments: exemplification.

I. RIVAL THEORIES OF CIRCULARITY

The simplest definition of 'circular argument' is "argument whose conclusion is also a premise." Proponents of the syntactic approach like the spirit of this definition because it portrays circularity as a relationship between the components of the argument. They favor analyzing circularity as a formal fallacy, such as affirming the consequent or undistributed middle; syntactic theorists say that whatever is bad about circular arguments is due to the *form* of the argument.

Despite their sympathy with the basic idea of the simple definition, syntactic theorists agree that it is too narrow. The definition only states a sufficient condition because it fastens on the most salient cases of circularity. Wild specimens of the fallacy conceal the repetition. Usually the identity of premise and conclusion is disguised by reclothing the conclusion in different terminology or grammar. Sometimes the conclusion is smuggled into the premise set by disassembling it into unfamiliar looking parts. Judicious juxtaposition with other premises also aids the illusion.

These complications are met with amendments aimed at outlawing these subterfuges.[1] This is as tricky as writing loopholes out of a tax code. To avoid condemning most valid deductions as circular, you must be specific. But it always seems to be just a matter of time before someone weasels through the clauses. When you then interpret the rules loosely to achieve a *de facto* elimination of loopholes, critics complain that you are exploiting the vagaries of propositional identity in an *ad hoc* fashion.

---

* An ancestor of this paper was read at the Eastern Division APA meeting in 1989. It has been improved by the reply of my commentator, Virginia Klenk. I also wish to thank Jonathan Adler, John Carroll, Julia Driver, David Sanford, Douglas Stalker, Kent Wilson, Allen Wood, and the editors of this journal for their comments and suggestions.

[1] For a recent example of a syntactic definition of 'circular argument', see p. 49 of Kent Wilson's "Circular Arguments," *Metaphilosophy*, XIX, 1 (January 1988): 38–51. Since Wilson distinguishes between circularity and vicious circularity, he would not regard himself as legislating against circularity. He does regard question-begging as always fallacious and so would reject my belief that 'circular' and 'question-begging' are synonymous.

0022-362X/91/8805/245–66

This mess led many syntactic theorists to develop dialogue games. The strategy is first to figure out which rules will preclude circularity in simple, formal dialogues and then to regulate progressively enriched language games.[2] The dialogue format introduces greater scope because disputation often involves relationships *between* whole arguments. Codification of rules and moves controls the complexity costs of the enhanced realism. The gradualism behind moving from simple to enriched games also strikes a nice compromise between piecemeal and systematic attacks on problems.

The pragmatic approach views circularity as a relationship between the argument and the people using it. For example, David Sanford[3] says that

> [Question-begging] is not a purely formal matter. An argument formulated for Smith's benefit, whether by Smith himself or by another, begs the question either if Smith believes one of the premises only because he already believes the conclusion or if Smith would believe one of the premises only if he already believed the conclusion. A primary purpose of argument is to increase the degree of reasonable confidence which one has in the truth of the conclusion. Not every argument which fails this purpose begs the question, but every question-begging argument fails this purpose (*ibid.*, p. 198).

This passage contains two sufficient conditions for circularity and a necessary condition, but not a condition that is both necessary and sufficient. The conditions featured in this partial analysis are broad, however, and accord well with views of other writers who have emphasized the epistemic aspects of inference.[4] Other pragmatists provide more detailed accounts of how circular arguments fail to be knowledge-expanding. William Lycan and George Schumm[5] say the problem is that the conclusion only differs trivially from the premises: circular arguments are valid but uninformative. They say that the vice of circularity is dialectic inefficiency. Frank Jackson[6] analyzes circularity as egocentric reasoning: the arguer selects premises in accordance with his own background beliefs rather than his audience's. Jackson also analyzes many apparently circular arguments as instances of "misleading advertising" because they have argumentative implicatures that mislead the audience about the nature of the arguer's evidence.

---

[2] C. L. Hamblin introduced the dialogue approach in *Fallacies* (London: Methuen, 1970). J. D. Mackenzie has refined the account in a series of articles, including "Why Do We Number Theorems," *Australasian Journal of Philosophy*, LVIII, 2 (June 1980): 135–49. Also see Frank Van Dun, "On the Modes of Opposition in the Formal Dialogues of P. Lorenzen," *Logique et Analyse*, LVII–LVIII (1972): 103–36; and Lauri Carlson's *Dialogue Games* (Dordrecht: Reidel, 1983).

[3] "Begging the Question," *Analysis*, XXXII, 6 (June 1972): 197–9.

[4] In particular, Sanford cites W. E. Johnson's *Logic*, Susan Stebbing's *A Modern Introduction to Logic*, and Max Black's "Induction and Experience," in *Experience and Theory*, Lawrence Foster and J W. Swanson, eds. (Amherst: Massachusetts UP, 1970).

[5] "A Theory of Question-Begging" (unpublished manuscript).

[6] *Conditionals* (New York: Blackwell, 1987), ch. 6.

Pragmatists do not agree with the spirit of the simple definition of circularity. Although they grant that the definition provides a simple sufficient condition for circularity, they lament the success of this condition on the grounds that it misleads people into thinking that circularity is a formal fallacy. Pragmatists think that the syntactic theorists have been mesmerized by the cases involving blatent repetition of the conclusion. The one-sided diet of examples fed to us by the sufficient condition encourages us to assimilate all cases of circularity to these unrepresentative "paradigms" of circularity. Compare this misguided reduction to our tendency to view all words as names. Pragmatists deny that any amount of tinkering with the simple definition will yield an adequate definition. They insist that charges of circularity must be relativized to audiences with varying purposes, background beliefs, and inferential prowess.

I am a gung-ho pragmatist. I think my brethren are too soft on the syntactic theorists. We should totally reject the simple definition that starts the syntactic ball rolling because propositional repetition is not a sufficient condition for circularity. Indeed, we should hit them where it really hurts. We should deny that all '*P*, therefore, *P*' arguments are circular. If the syntactic theorist cannot assume that '*P*, therefore, *P*' suffices for circularity, he has to close up shop; for this is the simplest case of all, the seed value for all his formulas.[7]

II. EXAMPLES THAT DO NOT WORK

A genuine counterexample to "All '*P*, therefore, *P*' arguments are circular" must express the same proposition in the premise and conclusion. We can achieve orthographic identity through devices of equivocation, such as polysemy (multiple senses), homonymy, and indexicality. But orthographic identity is not enough for propositional identity. Assume that in (A1) 'child' is ambiguous between 'offspring' and 'immature offspring' (respectively), that in (A2) 'ring' is homonymous between sounds and jewelry, and that in (A3) 'this' is self-referring.

---

[7] John Woods and Douglas Walton advocate an eclectic approach in "Arresting Circles in Formal Dialogues," *Journal of Philosophical Logic*, VII, 1 (February 1978): 73–90. The dialogue approach cannot retain its identity and accept my claims about the noncircularity of some '*P*, therefore, *P*' arguments The rules would have to become highly pragmatic so that the resulting system would look more like a web of Gricean maxims than a formal game.

(A1) Steven is Lori's child.
Steven is Lori's child.

(A2) John recalled two rings.
John recalled two rings.

(A3) This is the first token of this sentence.
This is the first token of this sentence.

These readings make all three of these arguments noncircular (and indeed invalid). They do not really have the '*P*, therefore, *P*' form, however.

III EXAMPLES THAT WORK

I now turn to examples that support my thesis that some '*P*, therefore, *P*' arguments are not circular. To indicate my judgment of their relative plausibility, I begin with the strongest and work down to the weakest.

A. Self-supporting versions. Some arguments are self-supporting in the sense that features other than validity buttress the conclusion:

(B1) If there are syllogisms, then information can be combined.
There are syllogisms.
Information can be combined.

(B2) An argument with a missing premise is an enthymeme.
There is an enthymeme on this page.

Since (B1) is an argument with two premises, it is a syllogism and hence evidence for its second premise, which in turn enhances the credibility of the conclusion. Argument (B2) is itself an enthymeme and so ensures the truth of its own conclusion. The following '*P*, therefore, *P*' argument is self-supportive in the same way:

(C1) Some arguments are written in black ink.
Some arguments are written in black ink.

In addition to be being sound, (C1) is rationally persuasive. A person who doubted the conclusion comes to know it is true by considering the argument. Since no rationally persuasive argument is circular (recall Sanford's necessary condition), (C1) is a noncircular '*P*, therefore, *P*' argument.

Self-support varies in kind and degree. Witness the mild self-help of

(C2) MORE THAN ONE ARGUMENT IS WRITTEN IN CAPITAL LETTERS.
MORE THAN ONE ARGUMENT IS WRITTEN IN CAPITAL LETTERS.

This argument token helps itself by filling half of its own quota but it does not fully confirm itself. Another parameter is the universality of self-support. We can capture this notion by defining self-support for argument types in terms of self-support for argument tokens. Call an argument *type* universally self-supporting exactly if all its tokens support themselves. Say that the type has limited self-support just in case only some of its tokens are self-supporting. The (C1) and (C2) argument types have limited self-support because there are green tokens of the (C1) type and "de-capitalized" tokens of the (C2) type. Examples of universal self-support can be obtained by exploiting formal properties of arguments:

(C3) Some arguments are composed solely of existential generalizations.
Some arguments are composed solely of existential generalizations.

(C4) It is possible for a conclusion to entail its premise.
It is possible for a conclusion to entail its premise.

(C5) Some deductive arguments do not reason from general to particular.
Some deductive arguments do not reason from general to particular.

Since (C3)–(C5) owe their persuasiveness to their logical form, they should be acceptable to even those with a narrow conception of logic. Arguments exploiting pragmatic properties make up an intermediate case:

(C6) Someone has thought about an argument.
Someone has thought about an argument.

(C7) An argument has been propounded by someone at some time.
An argument has been propounded by someone at some time.

Anyone who thinks about (C6) thereby confirms it. Since all tokens of the (C6) type are confirmed in this obvious fashion, it is a self-supporting argument type. If you propound (C7), it is also rationally persuasive. But if a (C7) token is merely mentioned, as when we consider an argument as an abstract entity, then it is not self-supporting. But since we tend to consider arguments as if they had been propounded, (C7) is apt to strike us as a self-supportive argument type.

Self-supporters can also be driven by properties of their *components:*

(D) Either no one can read or some people refuse to follow directions.
Someone can read.
Some people refuse to follow directions.

Reading the second premise confirms it and hence strengthens the argument. Now consider a '*P*, therefore, *P*' version:

(E1) Some sentences are exactly seven words long.
Some sentences are exactly seven words long.

In (E1), the real persuasive work is done by a property of the premise alone rather than by the argument. But when the statements are relationally self-supporting, both components play a role:

(E2) There are at least two tokens of an eleven-word sentence.
There are at least two tokens of an eleven-word sentence.

The premise and conclusion work together in (E2). Nevertheless, they would be just as persuasive if they had been merely listed. Since many people are more impressed by examples where the rational persuasion is done by an argument in its role as argument, they prefer the earlier examples.

Self-supporting '*P*, therefore, *P*' arguments are seldom propounded because we are rarely interested in establishing the sort of conclusions they demonstrate. But there are a few used in ordinary life. For example, I challenge your claim that clocks are to be set forward in spring and back in fall (as part of "daylight savings time"). You reply, "Spring forward, fall back."[8] The proposition expressed as your reason is identical to the conclusion. Nevertheless, the argument is persuasive because your premise is couched in a style typical of mnemonic sayings. My best explanation of your premise is that you are manifesting memory. For a second case, suppose you claim that Ronald Reagan phoned you last night. I confess that it was a practical joke: "It was only I imitating Ronald Reagan." You demand proof. I reply by repeating my confession in my Ronald Reagan voice. Similarly, I back 'I speak French' with 'Je parles francais'.

---

[8] J. D. Mackenzie reports persuading a challenger that April has only thirty days by replying, "Thirty days hath September, April, June, and November"—"Why Do We Number Theorems?" *Australasian Journal of Philosophy*, LVIII, 2 (June 1980): 135–49; p. 140. The form of this argument is '*P* & *Q* & *R* & *S*, therefore, *Q*' rather than '*P*, therefore, *P*', so I had to use a slight variant.

Grumps are leery of the (C) and (E) examples on the grounds that they involve self-reference. But this is 1991! 'Self-reference' is no longer a dirty word. Contemporary commentators on the liar paradox agree that Russellian knee-jerk rejection of self-reference would deprive us of substantial theoretical and practical benefits. Nevertheless, philosophers who view the liar paradox from a distance latch onto the self-referential element and do not let go. They know that *something* is deeply awry with 'This statement is false', they know that the liar statement is self-referential, and so come to have a vague but tenacious distrust of anything that smells of self-reference.

The assertions in the (C) arguments refer to properties of *arguments*, not assertions, and so are not about themselves. In the (E) cases, the assertions are about properties of assertions but do not refer to themselves with *singular terms*. The conclusions of the (E) (and (C)) arguments are existential generalizations and, although such statements are made true by individuals, they do not refer to them; 'Some men are divorced' does not refer to Johnny Carson. Sure, the (C) and (E) arguments do refer to themselves in the stretched sense of containing *predicates* whose extensions include the arguments or constituents of the argument. In this wide sense, 'There are visible things' is self-referential because it is visible. But once 'self-reference' is construed this broadly, it can hardly be a diagnostic term that picks out a property warranting suspicion.

The griping is really set off by simultaneous exemplification and denotation.[9] Each (C) and (E) argument (or its constituents) exemplifies the predicate its conclusion claims to be instantiated. Exemplification is narrower than the weak sort of "self-reference" discussed in the previous paragraph because exemplification is selective. Unlike the (C) and (E) arguments, the following does not exemplify the predicate asserted to be instantiated by the conclusion.

(F) Some arguments are typed left-handed.
―――――――――――――――――
Some arguments are typed left-handed.

In this case, the problem is inaccessibility: you cannot readily tell whether (F) instantiates 'typed left-handed', so (F) is not an exemplar. In contrast, the capitalization of (C2) is ostentatious.

---

[9] The relevance of Nelson Goodman's notion of exemplification was brought to my attention by Douglas Stalker. As my talk of propositions suggests, I have detached exemplification from Goodman's nominalism. For an extended discussion of exemplification, see Catherine Elgin's *With Reference to Reference* (Indianapolis: Hackett, 1983), ch. 5.

Exemplification is a subconverse of denotation: a denoting predicate refers to its object, while an exemplar is an object that refers to its predicate *and* instantiates it. So model homes, promotional handouts, and sample proofs all qualify as exemplars. Philosophically richer examples include Thomas Kuhn's paradigms, Iris Murdoch's moral exemplars, and the aesthetic works analyzed by Nelson Goodman. Some words simultaneously denote and exemplify, consider 'word', 'short', 'polysyllabic', and onomatopoeia such as 'bow-wow' and 'ding-dong'. The (C) and (E) arguments also appear to denote and exemplify simultaneously. For example, (C1) exemplifies the predicate 'argument written in black ink' used in the conclusion, while this predicate also denotes the argument. Normally, the interesting properties of an argument are what its premise and conclusion say. The (C) and (E) arguments show that the predicates exemplified can also be important. There is a parallel with poetry. It is natural to concentrate on what a poem says and to overlook what it exemplifies. As Goodman[10] observes, translation of poetry sometimes brings the omission to attention:

> The distinction illustrated here [between samples and labels] becomes important in the translation of poetry or other literature. Original and translation will of course differ in some properties; but so do any two inscriptions of the same word, or even any two red inscriptions of "red". The goal is maximal preservation of what the original *exemplifies* as well as of what it says. Translation of a staccato by a legato text may do greater violence than some discrepancies in denotation (*ibid.*, p. 60).

The point can be extended to argument translation. The natural Japanese translation of (E1) contains only six words, not seven. So, if this article were translated into Japanese, the sentence's description of sentence length should be revised to six words. Or suppose I wish to lecture on (C1) using a blackboard: 'white chalk' should be substituted for 'black ink'.

Self-referential worries are not the only obstacle to acceptance of the (C) and (E) cases as counterexamples to "All '*P*, therefore, *P*' arguments beg the question." One vague suspicion is that an argument such as (C1) *causes* belief in the conclusion without providing a *reason* for the conclusion. One can imagine a superneurologist causing you to believe that giraffes swim with the carefully calculated optical properties of the inscription 'Toads smile'. But his 'Toads smile, therefore, giraffes swim' argument would not give you reason for the conclusion even though it triggers belief in it. Although I agree that this story shows that there is a difference between merely causing belief and giving reason to believe, I deny that the distinction cuts against (C1) or any of the other examples. After all, they do ***rationally*** persuade. The cause/reason distinction is a red herring.

---

[10] *Languages of Art* (Indianapolis: Hackett, 1976), 2nd ed.

Appreciating the irrelevance of this distinction may lead one to draw another distinction that is relevant. 'Reason' sometimes refers to a justifying proposition and sometimes refers to things the proposition is about. Thus I may explain my hospitalization with the propositional reason, 'My leg broke', or by pointing to my broken leg as my "ontic" reason. We are used to thinking of arguments as aiming to have premises that are propositional reasons for their conclusions, but the (B)-(E) cases show that an argument itself can be an ontic reason for its conclusion. The population of actually uttered arguments falls into four categories:

	Ontically persuasive	Ontically unpersuasive
Propositionally persuasive	Few arguments	Most arguments
Propositionally unpersuasive	Very few arguments	Many arguments

An argument that is ontically but not propositionally persuasive will strike some as a "conflict case" between two conceptions of argumentation:

(G1) The moon is edible.
———————————
Some argument is written in black ink.

An objector might insist that we must *choose*. He will say that, for *logical* purposes at least, we should only consider propositional persuasiveness. So the objector concludes that the (C) and (E) arguments are circular after all: they rationally persuade in the wrong way (ontically).

Aversion to ontic persuasion is a symptom of validity chauvinism. If we read in a goal of perfect truth preservation, then (G1) is a bad argument in the sense that it fails to achieve the propounder's objective. But imputing this goal to inductions equally well condemns them as bad arguments. As in the case of induction, charity dictates the attribution of another goal when evaluating (G1). In particular, the propounder of (G1) can legitimately aim at persuasion by exemplification. Ditto for:

(G2) Some arguments are written in black ink.
Some arguments are written in English.
Some arguments are self-supporting in two ways.

(G3) If this argument is black, then it is the color of coal.
This argument is the color of coal.

(G4) Marilyn Monroe had six toes on her left foot.
Some arguments have a premise that is superfluous to its validity.
Some arguments have a premise that is superfluous to its validity.

Like (G1), (G2) and (G3) are rationally persuasive but invalid. (G4) is valid but this feature takes a back seat to the aim of persuasion by exemplification.

The validity chauvinist might grant that an arguer can have ontic persuasion as a goal and yet complain that this goal is an *illegitimate* aim. He might buttress this charge of logical perversion with an appeal to the frequency with which people balk at arguments such as (G1). But notice that students in logic class also complain of arguments that persuade in the wrong way. They balk at valid arguments corresponding to the paradoxes of strict implication. Students hunger for relevance and so resist sound arguments having unrelated tautologies as premise and conclusion. They want efficiency and so are ambivalent about the result of adding even "relevant" premises to a deduction previously agreed to be acceptable. Students demand credible premises and so must be dragged through their first reductios. These misguided balks should make us wary of resting on the balk against (G1). Granted, (G1) is invalid and has a premise that is both false and not topically related to the conclusion. But so what? It does rationally persuade in its peculiar way and logic is about rational persuasion through argument. Should logicians not explore this path rather than zone it off as outside the domain of logic? The burden of proof is on the balkers.

I see only two plausible ways for them to shoulder this burden. First, insist that (C) and (E) arguments are all *enthymemes;* and say the missing premises in each are self-referential propositions to the effect that the argument (or its components) instantiates the predicate the conclusion claims to be instantiated. But this proposal violates the maxim telling us to add premises only out of charity. The (C) and (E) arguments are already valid and the extra premises do not free them from the charge of circularity. Indeed, filling in missing premises makes some of them look worse:

(C3′) This argument is composed solely of existential generalizations.
Some arguments are composed solely of existential generalizations.
―――――――――――――――――――――――――――――――
Some arguments are composed solely of existential generalizations.

Whereas the original (C3) is sound and quite persuasive, (C3′) is unsound and less persuasive. Logician, do no harm!

The next objection appeals to meta-arguments rather than enthymemes: the (C) and (E) cases are instances of *indirect* rational persuasion. Thus (C1) only persuades us by bringing a meta-argument to our attention:

(C1.1) (C1) is an argument written in black ink.
―――――――――――――――――――――――――――――――
Some arguments are written in black ink.

Notice that (C1.1) does not have the '*P*, therefore, *P*' form but does have the same conclusion as (C1). The allegation is that, since (C1.1) is doing the ultimate work of rational persuasion, (C1) is really circular after all. (C1) presents an illusion of cogency because it resembles an argument that is cogent and that is immediately brought to mind by (C1). The resemblance and immediacy make it seem as if only (C1) is involved in the persuasion process when in fact there are two arguments.

My first reply is that the objection assumes that indirect persuasion is not persuasion. Compare indirect persuasion with indirect perception. When I see the wind blowing through the trees, I see the wind even though I am only seeing it indirectly through its effects.

My second reply is a charge of special pleading: to postulate a meta-argument as the ultimate persuader for (C1) is selectively to enforce a superficially plausible principle of justification. The basic version of the principle is that a conclusion is justified only if its premise and the link between the premise and the conclusion is justified by another argument. Conscientious application of the principle would not stop with the postulation of (C1.1). Other meta-arguments would have to be marshalled to justify the premise of (C1.1), for example

(C1.11) (C1) is in black print.
Most black print is composed of ink (rather than laser burn).
―――――――――――――――――――――――――――――――
(C1) is written in black ink.

After supplying arguments for the truth of (C1.1)'s premise, we would be obliged to supply arguments to establish its validity:

(C1.12) (C1.1) has the form of existential generalization.
Existential generalization is a valid argument form.
(C1.1) is a valid argument.

Now, at least, it seems we have satisfied the obligations imposed by the justification principle. One reward of our circumspection is that we see that even (C1.1) is only indirectly persuasive; for it depends on a cluster of background arguments needed to assure us that its premise is true and that the conclusion is truly implied. But on reflection, we see that our justificatory work has only begun! What about the soundness of the arguments supporting (C1.1)? Unless we accept them on blind faith, we will have to inquire into their logical credentials. And as long as we are going to do the job, let us do it right and inquire into subtler properties of arguments. We need to postulate further meta-arguments to reflect factors, such as the arguer's intentions, the conventions of disputation, and the argumentative implicatures conveyed by the selection and arrangement of premises and indicator words, such as 'since' and 'therefore'. Thus the metaargument objection is undermined by an infinite regress akin to Lewis Carroll's.[11]

B. Self-defeating versions. Whereas self-supporting arguments help themselves, self-defeating ones hurt themselves.

(H1) All arguments are written in green ink.
All arguments are written in green ink.

No circular argument is rationally dispersuasive. The complaint against circularity is that you are standing in place when you think you are going forward; not that you are going backward. Hence (H1) is a second example of '*P*, therefore, *P*' without circularity. Like self-supporters, self-defeaters can be affected by their formal properties:

(H2) No argument has a negative premise.
No argument has a negative premise.

(H3) All arguments are syllogisms.
All arguments are syllogisms.

Unlike (H1), both of these arguments are universally self-defeating.

The idea that you can weaken your case merely by arguing for it (or arguing for it in a particular way) is not esoteric. When Joe Sixpack is told that he is too defensive, contentious, or sophistical, he will worry that a rebuttal will confirm the criticism. If you tell Joe that he overuses the charge of circularity and you back up the accusation with the claim that he uses 'That's question-begging' as an argumentative crutch, he will be reluctant to object that your argument is circular. A variety of philosophical arguments have been alleged to be self-defeating. For each of the following, there are critics who say the position is undermined by the very act of arguing for it: behaviorism, cognitive pluralism, determinism, dogmatism, eliminative materialism, epiphenomenalism, epistemological anarchism, evolutionary epistemology, mechanism, nominalism, psychologism, solipsism. Ditto for arguments against free speech, objective truth, the meaningfulness philosophical discourse, and laws of logic. The charge of argumentative self-defeat is illustrated by Israel Scheffler's[12] critique of those who argue for the subjectivity of science on scientific grounds:

---

[11] "What the Tortoise Said to Achilles," *Mind*, IV, 14 (April 1895): 278–80.

> . . . there is a striking self-contradictoriness in the effort to persuade others by argument that communication, and hence argument is impossible; in appeal to the facts about observation in order to deny that commonly observable facts exist, in arguing from the hard realities of the history of science to the conclusion that reality is not discovered but made by the scientist. To accept these claims is to deny all force to the arguments brought forward for them (*ibid.*, p. 44).

Avoiding self-defeat is a ***standard problem*** for radical philosophical positions. This is evidence that appeals to self-defeat are part of accepted practice in argument evaluation and hence are of interest to logicians in the same way as appeals to validity.

No doubt, validity is far better understood than self-defeat. Much of the obscurity clears with careful definition and patient distinctions, however. Count an act as self-defeating just if it frustrates the achievement of its purposes. (Purposes are assigned to acts in the way they are to artifacts, such as pans, pens, and pins.) Thus it is self-defeating to struggle in quicksand and to use highbeams on foggy roads. Since an utterance is an act, we can infer that an utterance is self-defeating exactly when it frustrates the achievement of one of its purposes. Examples of linguistic backfire include:

1. The successor of the largest integer ever thought of is 192,394,031,331.
2. The fact that turtles gargle is ineffable.
3. It is possible that zebras hibernate but no one can conceive it.
4. Be spontaneous.
5. Never say 'never'

---

[12] *Science and Subjectivity* (Indianapolis: Bobbs-Merrill, 1967).

Correct application of 'self-defeat' requires knowledge of what the speaker was trying to do and what he has actually done in the attempt. Since the speaker has many purposes and his utterance instantiates many descriptions, 'self-defeat' should be relativized to these purposes and descriptions. The consequent relativity of self-defeat is illustrated by 'I never warn' which is self-defeating *as a warning* but not *as an assertion*. Utterance descriptions grow in complexity. Slowly saying 'I never speak slowly' is self-defeating, as is slowly saying in a Texas drawl 'I never talk in a slow Texas drawl', and slowly saying in a Texas drawl on Sunday 'I never talk in a slow Texas drawl on Sundays'. The larger utterances are not self-defeating, however, relative to the shorter utterance descriptions.

The heterogeneity of speaker goals can be gleaned from the variety of speech acts. 'Never take advice' is self-defeating because it cannot achieve the standard point of advising: being fruitfully heeded. A principal purpose of asking a question is to learn the answer from the hearer. This makes 'Can anything be learned by asking questions?' self-defeating. The very act of asking presupposes belief in an affirmative answer and hence precludes the possibility of achieving the goal of learning the answer. The head quote of Irving Copi's logic text contains an excerpt from the *Discourses of Epictetus* featuring a request that is self-defeating for much the same reason:

> When one of his audience said, "Convince me that logic is useful," he said,
>
> "Would you have me demonstrate it?"
>
> "Yes."
>
> "Well, then, must I not use demonstrative argument?"
>
> And, when the other agreed, he said, "How then shall you know if I impose upon you?" And when the man had no answer, he said, "You see how you yourself admit that logic is necessary, if without it you are not even able to learn this much—whether it is necessary or not."

Many speech acts (assertion, warning, promise) strive for plausibility, so a wide group of utterances can be *evidentially* self-defeating. The degree to which they are evidence against themselves varies from the certain self-defeat of 'There are no sentences' to the mildly self-defeating twelve-word sentence, 'At most one sentence on this page is exactly twelve words long'. A second large group of speech acts aim at compliance—a world-to-word fit instead of a word-to-world fit. So we can say an utterance is *normatively* self-defeating when it constitutes an infraction of the imperative it expresses. Unlike 'No sentence is six words long', which counterexamples the proposition it expresses, 'Never write a six-word sentence' disobeys the imperative it expresses. Graffiti writers who scrawl 'Do not write on the wall' on the wall wittingly violate the prohibition but do not falsify it. Archibald MacLeish unwittingly violates his proscription against didactic poetry when he ends his poem "Ars Poetica" with:

> A poem should be equal to:
> Not true
> For all the history of grief
> An empty doorway and a maple leaf
> For love
> The leaning grasses and two lights about the sea—
> A poem should not mean
> But be

Notice that MacLeish's utterance is normatively self-defeating only as a poetic verse, not as a simple assertion or as the thesis of an argument. The relativity of self-defeat is also illustrated by Elmore Leonard's *52 Pick-up*. In the novel, a blackmail victim learns that his gun killed his mistress by viewing a video tape that was filmed in a way that eliminated the possibility of a fake execution. This is not a self-defeating depiction but the movie version of *52 Pick-up* is; for one scene shows the video tape, which at first does appear unfakeable except when you reflect on the fact that Hollywood does not make snuff films.

In most cases of self-defeat, the infraction is by the speaker. There is, however, a genre of self-defeat in which the audience must break the rule in order to understand it, as with 'Do not read this sentence' and 'Do not think of pink elephants'. Audience participation is also crucial to self-defeating prophecies in which dissemination of the prediction causes a reaction that falsifies it. For instance, the agricultural expert's prediction of corn shortages may prompt farmers to plant more corn.

A "self-defeating argument" can be defined as one that contains a self-defeating premise or conclusion. This allows some arguments to be self-defeating in ways that are independent of their status as argument. For example, an argument for the conclusion, 'No sentence is six words long', contains a self-defeating *assertion*. Incorporating 'There are no arguments', however, yields a self-defeating *argument* even though the statement is not self-defeating as a bare *assertion*. This pure sort of argumentative self-defeat can be normative. Robert Nozick's[18] argument against arguing is normatively self-defeating only because it is an argument; as a bare protest, it does not frustrate itself.

---

[18] *Philosophical Explanations* (Cambridge: Harvard, 1981), pp. 4–5.

The analysis of self-support parallels that of self-defeat; merely substitute 'fulfill' for 'frustrate'. This extension is important because exemplification is openly intended self-support. When an applicant for the job of translator says 'I speak English' to a skeptical interviewer, he intends his assertion as a sample. Ditto for the salesman who argues that he ought to be hired because he is a persuasive speaker, the testifier fending off accusations of insanity, and the logician demonstrating that the study of logic reduces your vulnerability to fallacies. Clever refutations often exemplify what they assert. My favorite was aimed at a linguistic philosopher who was delivering a lecture in which he just noted that the double negative construction in some languages has a positive meaning. He went on to note that, in other languages, the double negative intensifies denial. Yet, in no language, said the lecturer, do we find a double positive construction with a negative meaning. From the back of the lecture room, someone interjected in a jeering voice "Yeah, yeah." A more famous example is Winston Churchill's objection to the prohibition against ending sentences with prepositions: "That is something up with which I will not put." The objection is more than a bare protest because it exemplifies the awkwardness compliance would cause.

Exemplification is at play in a number of philosophical arguments. In the cogito, Descartes's "mental utterance" of 'I think, therefore, I exist' exemplifies his thinking, thereby establishing the truth of the premise. The exemplification feature provides a reply to the objection that Descartes could have just as well used 'I walk', 'I eat', or some other less ponderous assertion as his premise.

More obscure examples can be drawn from a project undertaken by Girolamo Saccheri[14] in 1697. Saccheri aimed to prove the invalidity of invalid syllogisms using rules designed to prove the validity of valid syllogisms. His method can be illustrated with his argument against AEE-1 syllogisms, also called "Maltese" syllogisms.

All Barbara syllogisms are valid.
No Maltese syllogisms are Barbara syllogisms.

No Maltese syllogisms are valid.

---

[14] Saccherian syllogisms are discussed in William and Martha Kneale, *The Development of Logic* (New York: Oxford, 1962), pp. 345–8.

The premises are true. So, if the argument is valid, the conclusion must be true. But since the argument is a Maltese syllogism, it would then follow that the argument is invalid. Therefore, Maltese syllogisms are invalid—just as the conclusion asserts. The fascinating feature of Saccheri's argument is that one must recognize its (formal) invalidity in order to be persuaded by it. So it should confound the validity chauvinist.

Many logicians would be leery of Saccheri's syllogism's apparent self-referentiality.[15] But the conclusion does not imply that the argument is itself invalid. That verdict only follows with the ancilliary premise that the argument instantiates the Maltese form. Although this is never *asserted* in the argument, the argument does *exemplify* 'Maltese syllogism'. So the notion of exemplification helps to extricate Saccheri's syllogism from the objection that it embodies a confusion between language and metalanguage.

C. Arguments having logical theorems as their conclusions. My third group of examples is inspired by the premiseless arguments discussed in logic textbooks. The typical conclusion can only be proved by applying inference rules to premises. But if the conclusion is a theorem of logic, the inference rules alone are enough to do the job. Adding a premise to a valid premiseless argument yields a valid single-premise argument whose premise is superfluous. Since we are free to choose any premise, let us take the conclusion itself as the premise, as in:

(I) Either ants spit or it is not the case that ants spit.
Either ants spit or it is not the case that ants spit.

Given that the conclusion is inferred purely by means of inference rules, there is no self-inference, and so no circularity. Yet the argument has the form '*P*, therefore, *P*'.

For (I) to work, it must be understood as an argument with a single, redundant premise. Relative to a solely rule-driven derivation, (I) is not circular. Yet (I) is circular relative to a derivation that makes essential appeal to the premise of (I). As is common practice in logic textbooks, we should distinguish between arguments and proofs. Arguments are not identical to proofs because one argument usually has many alternative proofs of its conclusion.

---

[15] C. L Hamblin defends Saccherian arguments in "Saccherian Arguments and the Self-Application of Logic," *Australasian Journal of Philosophy*, LIII, 2 (August 1975): 157–64

D. Examples drawn from pragmatics. Progress in pragmatics (indexicals, implicature, emotive meaning) has overturned the belief that utterances that express the same proposition must have the same meaning and utterances that have the same meaning must express the same propositions. For example, we distinguish between the occasion meaning of 'I bowl' and its constant meaning. Its occasion meaning is the proposition expressed on a particular occasion when we "fill in" the referent of 'I'. When Carter utters 'I bowl', its occasion meaning is that Carter bowls. When Bush utters 'I bowl', its occasion meaning is that Bush bowls. But it also has a meaning that does not vary from speaker to speaker. This constant meaning is not a proposition. Hence, there is a sense in which two statements that do not express the same proposition can mean the same thing. This suggests the possibility of circularity without propositional repetition. For example, pragmatic theories of sarcasm deny that it alters the truth conditions of what is said. The difference it makes to meaning consists of what the speaker suggests. This lends support to the intuitive circularity of

(J) There is a slim chance that the Mets will win.
―――――――――――――――――――――――――
There is a fat chance that the Mets will win.

Given that the premise is literal while the conclusion is sarcastic, they express incompatible propositions. Yet they mean the same thing in the sense of conveying doubt that the Mets will win.

In the case of constant meaning, we say that two assertions mean the same thing even though different propositions are expressed. Noncircular '*P*, therefore, *P*' arguments can be generated from the converse case, where we say that two assertions have different meanings even though they express the same proposition. A necessary condition for circularity is that the conclusion repeat something meant in the premise set. So, if the conclusion means something not contained in the premises, the argument is not circular. Once we allow assertions to mean more than the propositions they express, we allow for the possibility that a '*P*, therefore, *P*' argument will have a conclusion that means something different from its premise; for remember that *P* stands for a *proposition* and propositions do not exhaust meaning.

Grice's theory of implicature provides widely accepted evidence that meaning goes beyond the truth conditions of the statement. For instance, the premise and conclusion of the following express the same proposition but have different conversational implicatures:

(K1) Princess Di got married and pregnant.
Princess Di got pregnant and married.

Whereas the premise suggests that Di first married and then later conceived, the conclusion suggests the reverse temporal order. This will lead many of Princess Di's champions to affirm the premise but deny the conclusion. Even defenders who grant that the premise and conclusion express the same proposition will object that the argument is misleading—not that it is circular. Of course, those who hold that '*P* & *Q*' is a different proposition than '*Q* & *P*' will deny that (K1) has the '*P*, therefore, *P*' form. But such fastidiousness about propositional identity would be too picky for anyone who believed that circular reasoning is common in ordinary argumentation; for naturally occurring arguments usually have slight differences between premises and conclusions that "really say the same thing."

The conclusions of the next pair share their premise's truth conditions but differ in pragmatic meaning by an extraconventional implicature.

(K2) Mary is poor and honest.
Mary is poor but honest.

(K3) Tim can answer it.
Even Tim can answer it.

The conclusion of (K2) suggests that poverty tends to preclude honesty. The conclusion of (K3) gives the impression that the question is easy. Once again, I deny that these arguments are circular on the grounds that the premise and conclusions do not mean the same thing.

The divergence between semantic and pragmatic synonymy can be subtle. Highly status-conscious languages such as Japanese extend honorific usage beyond people to their groups, actions, and belongings. English has a pale shadow of this phenomenon in the form of elevation: (residence, home), (dine, eat), (lady, woman), (conjecture, guess), (postulate, assume), (perspire, sweat), (imbibe, drink). The left-hand member of each pair confers higher status than the right without any descriptive supplement. This difference between semantic and pragmatic synonymy is the basis for

(L1) Vernon owns a violin.
Vernon owns a fiddle.

Emotivists were attracted to more strident phenomena: honorific and pejorative terms. According to the emotivists, some expressions that are descriptively synonymous vary in emotive meaning.

(L2) Mary is pig-headed.
Mary is resolute.

(L3) Joe is a German.
Joe is a Kraut.

In (L2), we move from pejorative to honorific. In (L3), the premise lacks emotive meaning but is descriptively equivalent to its charged conclusion.[16]

Given pragmatic theories of emphasis and focus, these also work:

(M1) Hester is the winner.
Hester is *the* winner.

(M2) Good food is not cheap food.
Cheap food is not good food.

Argument (M1) brings the uniqueness condition of definite descriptions to the fore by emphasis. In (M2), the premise brings to mind good, expensive food, while the conclusion prompts thoughts of cheap, bad food. Once again, disagreement about propositional identity may lead some to deny (M2) has the '*P*, therefore, *P*' form on the grounds that converses are different propositions rather than different expressions of the same proposition.

The last examples are inspired by pragmatic theories of metaphor.[17] These theories say that metaphorical statements do not differ in truth conditions from their literal counterparts. Read the premise of the (M3) literally and the conclusion metaphorically, and do the reverse for (M4).

(M3) Tons of fish were in the boat.
Tons of fish were in the boat.

(M4) Mayor Koch has a big mouth.
Mayor Koch has a big mouth.

One can imagine the premise of (M3) used to justify the metaphorical conclusion. A person with a penchant for "resemblance thinking" might use the metaphorical premise of (M4) to justify his anatomical conclusion.

E. Aspect cases. Perhaps propositions are like gems. Just as the same gem can have many interesting facets, a proposition might have different aspects that vary in salience. Bearing this analogy in mind, read the following as an inference from a passive to an active change:

---

[16] J. O. Urmson has a good discussion of emotive meaning in *The Emotive Theory of Ethics* (London: Hutchenson, 1968), ch. 10

[17] Donald Davidson defends a pragmatic theory of metaphor in "What Metaphors Mean," in *On Metaphor*, Sheldon Sacks, ed. (Chicago: University Press, 1979), pp. 29–45. James D. Carney also defends it in "The Meaning of Metaphor," *Philosophical Studies*, XLIV, 2 (September 1983): 257–67.

(N1) The newspaper acquired new owners.
New owners acquired the newspaper.

Although the premise and conclusion have the same logical form, the premise highlights the newspaper's continuity. The slight orthographic difference between the premise and the conclusion might lead some to suppose that different propositions are involved. This defect is eliminated by the following example, which should be read as an inference from the existence of tenants to the existence of landlords:

(N2) Someone rents a house.
Someone rents a house.

If distinct propositions are involved, then the example is as bogus as the ambiguous ones discussed in the second section. Yet note that the premise and conclusion are orthographically identical and strictly entail each other. The felt difference between the readings could be explained as one of aspect. People attend to the existence of tenants when they first hear 'Someone rents a house' because tenants are more common than landlords. But the proposition also entails the existence of a landlord. The difference in salience makes belief in the landlord lag behind belief in the tenant. But redundancy prompts a second look, allowing the laggard to catch up.

IV ARE 'CIRCULAR' AND 'QUESTION-BEGGING' SYNONYMOUS?

There are at least two commentators who would object to the fact that I use 'circularity' and 'question-begging' interchangeably. Robert Fogelin's[18] logic text contains the following passage:

> Closely related to the fallacy of circular reasoning is the subtle fallacy called *begging the question*. In circular reasoning, the conclusion itself or a restatement of the conclusion in different words is used as a premise in the argument. In begging the question, a statement is made that presupposes or depends upon the point at issue (*ibid.*, p. 95).

Other writers, such as Kent Wilson, reserve 'question-begging' for arguments that are *viciously* circular. Although the synonymy deniers differ on how the meanings diverge, they will be tempted to cast my dispute with the syntactic theorist as merely verbal; for one might craft the distinction to reach the diplomatic solution that my '*P*, therefore, *P*' arguments are circular but not question-begging. 'Circular' might be defined as a purely syntactic concept and 'question-begging' as a pragmatic or semantic concept catching the residual nonsyntactic sorts of dependencies.

---

[18] *Understanding Arguments* (New York: Harcourt Brace & Jovanovich, 1987).

Etymologists tell us that synonyms are unstable; one synonym is either dropped or changes its meaning to cover a nuance. My opinion is that Fogelin and Wilson are pressuring 'circular' and 'question-begging' into the latter fate. In any case, the synonymy issue is a red herring; for even if the debaters grant that there is a difference in meaning, they will only reformulate the debate in the new terminology. The syntactic theorist will deny that he is only analyzing the diplomat's concept of circularity. His goal was to achieve a formal understanding or explanation of an *interesting* fallacy. Confining the meaning of 'circular' in a way that obviously guarantees the adequacy of a syntactic account attains tranquility at the price of triviality. So, if the syntactic theorist accepts the distinction, he will reformulate his thesis so that his enterprise encompasses question-begging. On the other hand, if 'circular' is not defined in a way that makes him obviously right, we lose confidence that the dispute is verbal. So my reply to the objection that 'circular' and 'question-begging' are not synonymous is (a) *reportive* definitions will show their synonymy and (b), even if they diverge, the difference will be irrelevant to my thesis.

V. CONCLUDING REMARKS

The noncircular '*P*, therefore, *P*' arguments are intended to be mutually supportive even though they are logically independent. Once you accept one example, resistance is broken and you are apt to accept many more examples.

The examples are more than curios. First, they are relevant to a controversy over how circularity is to be analyzed. Noncircular '*P*, therefore, *P*' arguments refute the traditional syntactic account of circularity and confirm the pragmatic approach in a way unanticipated by the pragmatists. Second, some of these examples suggest that there is a neglected parameter of evaluation: exemplification. Those who accept the suggestion will greet the cases as more than logic teasers.

ROY A SORENSEN

New York University

BOOK REVIEWS

History, Labour and Freedom: Themes from Marx. GERALD A. COHEN. New York: Oxford University Press, 1988. xiii, 317 p. Cloth $65.00, paper $24.95.*

History, Labour and Freedom (HLF) is comprised of fourteen essays by G. A. Cohen, based on papers published, with two exceptions, after the appearance of *Karl Marx's Theory of History: A Defence* (KMTH).[1] Six essays have been extensively revised or reorganized, four are reprinted virtually unchanged, and four have been modified slightly. Five of the papers in this collection continue the project begun in KMTH: the exposition and defense of historical materialism. There are also four essays in which Cohen himself takes aim at the theory defended in KMTH. Of the five remaining pieces, three deal with freedom; one with exploitation and the labor theory of value; and one, "The Dialectic of Labour in Marx," originally published fifteen years ago, grapples with notions of alienation and dialectics, themes largely ignored in analytical Marxian writings of more recent vintage. Cohen's papers on self-ownership and equality are not represented in this collection; nor is there much indication of his long-standing engagement with libertarian social philosophy.

Because it collects and updates Cohen's recent work on historical materialism, HLF is a convenient and timely addition to the literature KMTH has spawned. Cohen's position in this on-going discussion has become increasingly ambivalent. By his own account, he continues to defend the cogency of the theory developed in KMTH, at the same time that he has become skeptical of its truth. An uncharitable reader might conclude that he defends his reconstruction of the theory against all critics, while advancing hesitations of his own about its plausibility and importance. But however Cohen's stance is construed, the cogency of the arguments he adduces makes his recent positions, pro and con, as nearly indispensable to this debate as KMTH itself.

HLF is also useful for drawing together some of Cohen's contributions to the analysis of freedom, his critique of the labor theory of value, and, in the essay on the dialectic of labor, his account of Marx's intellectual debt to Hegelian philosophy. But for readers

---

* I am grateful to G. A. Cohen, Daniel Hausman, Debra Satz, and Erik Olin Wright for comments on an earlier draft of this review.

[1] Princeton: University Press, 1978. Published simultaneously in Great Britain by Oxford University Press.

0022-362X/91/8805/267-75

interested in Cohen's work on normative issues arising out of his penchant for taking libertarianism seriously, this collection will only whet the appetite. There is need for a second volume of essays responsive to this interest.

As in KMTH, Cohen's reconstructions and critical assessments of Marx's views consistently conform to the highest standards of argumentation and scholarship. Many positions of varying degrees of contentiousness are staked out and defended with great acuity. These essays therefore invite engagement informed by the conviction that, in philosophy and scholarship, "God is in the details." Inevitably, there are, in the details, ample opportunities to quarrel with Cohen's views. But there are also general themes that emerge, particularly in the essays on historical materialism, that call for consideration in their own right. At this point in the development of the debate KMTH originated, it is instructive to address Cohen's recent work at this level. I shall therefore focus on one of the themes advanced in this volume: the claim that historical materialism, if sustainable, explains less than Marxists have traditionally supposed, and the related idea that the theory matters very little for other Marxian theoretical endeavors or for Marxian politics. I shall take issue with Cohen on both counts: suggesting, first, that the explanatory pretensions of Marx's theory be diminished even more radically than Cohen proposes, but then insisting that, even so reduced, historical materialism plays a more important role in the Marxian project than Cohen maintains.

I

According to the theory reconstructed and defended in KMTH, human history—from the beginnings of class society until the advent of communism—has a determinate structure and direction, in consequence of an endogenous dynamic process in which social "relations of production" rise and fall in order to maintain functional compatibility with developing material "forces of production." Marx sketched this position in a few brief paragraphs in the 1859 Preface to *The Critique of Political Economy,* a text that has become canonical for "orthodox" historical materialists. No one since Marx has done more than Cohen—in KMTH and in some of the essays in HLF—to elaborate and develop the general schema announced in 1859: explaining what forces and relations of production are, how Marx conceived their interaction, and why forces tend to develop continuously, while sets of production relations change discontinuously to accommodate ever increasing levels of development of productive forces.

It might appear that the internal dynamic Marx postulated would seldom, if ever, figure in explanations proffered by practicing historians, because a historian's ***explanandum*** is typically more fine-grained than a historical materialist's. Marx's theory of history does not, after all, purport to explain *events*. It is a theory of history itself, of its (epochal) structures and of the processes that impel movement from one structure to another. Mainstream historiography assumes that history in this sense has no theory: that historical events, no matter how individuated, may all be explicable at least in principle, but that even in principle there exists no explanation for history itself—except perhaps in the trivial sense achieved by conjoining all explanations of particular events. For the explanatory objectives of practicing historians, it would therefore seem that Marx's theory of history would be of little interest. Arguably, the apparent irrelevance of historical materialism for Marx's own historical writings attests to the soundness of this inference.

When Cohen suggests that the Marxian project can survive a weakening of historical materialism's explanatory pretensions or even outright abandonment of historical materialism itself, his rationale does not directly appeal to historical materialism's ostensibly superfluous role in historical explanations, but to its irrelevance for normative criticisms of capitalist society. He writes:

> Some people may think that, in raising a challenge to historical materialism, I am letting the side down. So let me emphasize that my belated reservations about the theory do not weaken my belief that it is both desirable and possible to extinguish existing capitalist social relations and to reorganize society on a just and humane basis. The political significance of retreat from historical materialism should not be exaggerated. An appreciation of the principal evils of capitalism, which are its injustice, its hostility to the development of the faculties of the individual, and its voracious ravaging of the natural and built environment, does not depend on ambitious theses about the whole of human history. Nor does the claim that it is possible to establish a society without exploitation which is hospitable to human fulfillment require, or perhaps even follow from, those theses. So scepticism about historical materialism should leave the socialist project more or less where it would otherwise be (132).

I would suggest, however, that more is at stake in debates about Marx's theory of history than Cohen suggests in this passage; and that failure to appreciate historical materialism's foundational role in Marxian theory and politics can lead to an unnecessarily hasty rejection of Marxism itself.

The historical materialist orthodoxy Cohen reconstructed and defended in KMTH posits functional relations between forces and relations of production, and between sets of production relations or "economic structures" and legal and political "superstructures" and forms of consciousness. The most trenchant of the reproaches Cohen now levels against the theory of KMTH or, more precisely, against the traditional understanding of its explanatory scope, pertains to the second of these functional relations. On the traditional "inclusivist" view, economic structures explain everything—at some suitable (but unspecified) level of generality. What Cohen now suggests is that historical materialists need only hold that economic structures explain superstructural facts to the extent that is necessary to maintain the autonomy of the endogenous dynamic process that causes economic structures to rise and fall. For a "restricted historical materialist," noneconomic facts cannot impinge on this underlying process in the wrong way; they cannot upset its dynamic. But they can fail to affect that process at all. Insofar as they remain causally aloof, historical materialists have no reason to maintain that they are in any way explained by the process the theory identifies. Orthodox historical materialism was a theory of general history—that is, of virtually everything historical. In suggesting that Marxists have reason to be restricted, not inclusive, historical materialists, Cohen, while enhancing its plausibility, radically deflates historical materialism's explanatory pretensions. Restricted historical materialism would hardly count as a theory of general history. But it is a foundational theory nevertheless.

There is a second reproach Cohen levels against historical materialist orthodoxy that addresses the underlying dynamic it purports to identify. That process is propelled by a transhistorical human interest in developing productive forces. In several of the essays in this volume, Cohen defends this claim against charges of circularity. But then he advances other reasons for disputing its putative role in moving history along. He suggests that the historical dynamic may be shaped as well by other interests—particularly, as Hegel believed, by an interest in "self-identification." If so, historical materialism may fail not just as a theory of general history, but also as a general theory of history—that is, of historical trends—precisely because it is fatally one-sided.

Defenders of historical materialism can resist this conclusion, however, and still concede that an interest in developing productive forces is not the only properly transhistorical interest shaping human mentalities. Historical materialists can even concede the existence of a dialectic of self-identification of the Hegelian sort and invest it with historical efficacy, as Cohen suggests. To see why, it will help to follow Cohen's own lead by focusing just on what the theory can plausibly be thought to explain, regardless of the explanatory ambitions its adherents have traditionally assumed.

The historical materialism of KMTH purports to provide an account of necessary and sufficient conditions for epochal historical transformations—from one set of production relations to another. It holds that certain levels of development of productive forces are necessary for the inception and reproduction of particular "economic structures" or sets of production relations and that continuing development generates "contradictions" (structural instabilities) between forces and already existing relations of production. Thus, in the orthodox view, the development of productive forces is sufficient in the long run for ensuring movement from one set of production relations to another. It is only when productive forces are massively developed and harmony between forces and relations of production is ensured under socialism that this process finally ceases to govern the transformation of economic structures.

It is not immediately apparent what would count as evidence for or against the existence of this internal dynamic. But on a superficial reading of human history, it would seem to explain little, if anything, of what has actually transpired. The historical-materialist dynamic should figure prominently, for example, in explanations for the emergence of capitalism. But in England and the parts of Western Europe where capitalism first became the dominant economic structure—the one instance in which it was not introduced by (imitative) domestic elites or imposed by outside powers—even some Marxian historians accord explanatory priority to such factors as peculiarities of class structure, the pattern of town-countryside relations, the nature of political institutions, and demographic considerations. The fettering of productive forces under feudal relations of production does not seem to play an important explanatory role.[2] It would therefore seem that the orthodox account of necessary and sufficient conditions for change, like the orthodox view of the relation between base and superstructure, is embarrassingly susceptible to disconfirmation. The idea that the theory is misleadingly one-sided would explain this unhappy situation. But in the apparent absence of an analogue to restricted historical materialism to which to retreat, the implication is not that the theory be weakened in its explanatory ambitions, but that it be overthrown altogether. This appears to be the conclusion toward which Cohen now tends. If Marx was wrong to invest the dynamic of history exclusively in an interest in developing productive forces, the theory that follows from this premise would be flawed at its core. No wonder, then, that it should square poorly with the facts of the matter. But, *pace* Cohen, it is not necessary to destroy the theory in order to save it. There are less radical ways to shelter Marx's account of history's underlying dynamic—or, more precisely, the core insight that motivates it—from the likelihood of empirical confutation.

---

[2] Not surprisingly, debates about the origins of capitalism have long preoccupied Marxists. Several decades ago, this debate was waged in expressly Marxian terms in the pages of *Science and Society* in a series of articles sparked by Paul Sweezy's criticisms of Maurice Dobb's *Studies in the Development of Capitalism* (New York: International, 1947, repr. 1963, 1972). The *Science and Society* debate has been republished, with additional material, as *The Transition from Feudalism to Capitalism* (New York: International, 1976). Since the publication of Robert Brenner's article, "Agrarian Class Structure and Economic Development in Pre-Industrial Europe," *Past and Present*, LXX (February 1976): 30–75, the Marxian debate over the transition from feudalism to capitalism has been

One could, for instance, rely on the distinction between the general laws historical materialism advances and the hypothesis that these laws actually do play a prominent role in explaining real history.[3] If the case for historical materialism developed in KMTH is cogent, then there is good reason to believe in the reality of an endogenous process capable of moving history from one epochal structure to another. In the absence of other pertinent causal factors, this process would explain what actually transpires. But, of course, other factors are always present. Thus, those who account for, say, the emergence of capitalism by appeal to considerations distinct from the ones historical materialism postulates need only maintain that the causes they invoke swamp the historical-materialist dynamic; not that the process historical materialists identify does not exist. Then the theory of KMTH would be explanatory, but much less so than Marxists have traditionally supposed.

There is, however, a more plausible way to regard the Marxian theory of history which recalls the retreat Cohen himself proposed

---

joined by historians of a variety of theoretical orientations—in the form of competing explanations for different levels of economic performance in different societies in early modern Europe. Some of the principal contributions to this debate are collected in T. H. Aston and C. H. E. Philpin, eds., *The Brenner Debate: Agrarian Class Structure and Economic Development in Pre-Industrial Europe* (New York: Cambridge, 1985).

[3] Cf. Andrew Levine and Elliott Sober, "What's Historical about Historical Materialism?" this JOURNAL, LXXXII, 6 (June 1985): 304–26, esp. pp. 312–4.

in distinguishing restricted from inclusive historical materialism. This proposal ill accords with Marx's own understanding of the theory he developed or with the orthodoxy reconstructed in KMTH. But it does advance Marx's "rational intuition" in a plausible way—at the same time that it reveals, as perspicuously as even the most stringent orthodoxy, historical materialism's foundational role in Marxian theory. I have elsewhere called the sort of theory I have in mind *weak historical materialism.*[4] Weak historical materialism does not advance sufficient conditions for epochal historical change, but only necessary material conditions. It does not purport to explain what actually happens, other things being equal (or nonexistent), but only what is (materially) possible. By identifying correspondences and "contradictions" between forces and relations of production, it specifies a map of historically possible epochal structures, and also the possibilities for moving from one epochal structure to another. In short, it establishes a determinate *historical agenda.*

II

Whether in its weak or traditional form, historical materialism identifies social classes as the agents who move history along. In the Marxian scheme, social relations of production constitute class positions into which individuals fit; and it is the concerted activities of individuals, as class actors, who reproduce—or overthrow and establish—the economic structures that comprise history's epochal divisions. Thus, historical materialism forces attention on class structure, class relations, and class struggle. To be sure, proponents of weak historical materialism—particularly if they are also restricted historical materialists—are bound to endorse class analysis only insofar as the phenomena they seek to explain bear on the historical-materialist dynamic. But this commitment is enough to motivate a distinctive research agenda.

Understood in the usual way, as a claim about what explains particular phenomena, it is almost certainly false that class is always more explanatory than other factors. But the usual understanding is misleading. What drives the Marxian worldview is the idea that power—at the level of social structure—accounts for social order and change; and that real (as opposed to merely juridical) ownership of productive resources, the basis for class divisions in the Marxian sense, is the source of this power. Explicit class identifications or overt class conflicts need not have more explanatory force in particular instances than other considerations. But at the level of generality appropriate for understanding epochal change, class relations render social life intelligible. For Marx, intelligibility at this level is indispensable for a political practice directed toward the overthrow of capitalism and its replacement by socialism, its historical alternative and successor.

---

[4] See *ibid.*, pp. 323–6, and my *The End of the State* (London: Verso, 1987), ch. 5

Thus, despite Cohen's imperturbability, Marxists cannot so easily do without Marx's theory of history. The normative critique of capitalism Marxists proffer would indeed survive historical materialism's demise. But if moral criticism were the only theoretical resource propelling Marxists' struggles for socialism, Marxism would devolve into a utopian socialism of the kind against which Marx vehemently inveighed. Marxists have insisted repeatedly that their socialism is "scientific," not utopian, precisely because it is grounded in an account of real, material processes—not moral ideals, however compelling. Historical materialism, even in its restricted and weak form, is crucial for sustaining this claim. Marxism is not a utopian socialism precisely because its theoretical and political project is grounded in a theory of history's structure and direction.[5]

No doubt, Marx and his successors were carried away by their polemics against utopian socialism. They had no need to oppose, as they sometimes did, the very idea of normative theory. Indeed, Cohen has shown—no where more persuasively than in an essay included in this volume, "Historical Inevitability and Revolutionary Agency"—that orthodox historical materialism is compatible with and even requires moral criticism to motivate revolutionary agents. If historical materialism, properly revised, is not a theory of historical inevitability after all, the need for a clear perception of the evils of capitalism becomes all the more pressing. It is therefore urgent that socialists take back on board the moral imperative driving the utopian-socialist vision. Many of the essays in this collection contribute importantly to this task. But, in the final analysis, for just the reasons Marx adduced, it is not enough that socialism be morally necessitated. Marx insisted that socialism be inscribed in a historical dynamic. The historical materialism of KMTH, if sustainable, would ensure that it is. But it is not clear why "scientific" (nonutopian) socialist theory must endorse socialism's inevitability. It should be enough to defend its place on the historical agenda. This is precisely what weak historical materialism does.

---

[5] There, of course, remains the possibility that some rival account of history's structure and direction could do the work historical materialism does in the Marxian scheme. Perhaps, as Cohen suggests, a "dialectic of self-identification" might play some role in such a theory. At the present time, however, this proposal is only programmatic. Historical materialism is surely not the only imaginable theory capable of grounding the theoretical and political project Marx envisioned, but it is, for now, the only full-fledged theory extant.

The question of the viability of historical materialism—or of successor theories that retain its core insights—should therefore be an issue of great urgency for anyone concerned with Marxism. In moving from inclusive to restricted and then from traditional to weak historical materialism, the plausibility of the theory is enhanced, without detriment to its foundational role in class analysis and class politics. But is what remains true? No considerations so far adduced by Cohen or anyone else require us to conclude that it is not. But neither, at this point, can we be sure that the core insight historical materialism articulates can finally be sustained.

Whatever his current hesitations, Cohen remains the premier defender of Marx's theory. This is why his suggestion that historical materialism may ultimately prove false cannot fail to disturb. I think Cohen would agree that the jury is still out. He should also agree that the issue cannot be settled by philosophical argument alone. Since the aim of the theory is to make sense of history at a certain level of generality, historical evidence and judgment must ultimately adjudicate the theory's claims. But philosophers can help by clarifying what corroboration or disconfirmation involves. Cohen's recent work—even at its most skeptical—cogently addresses this complex and still scarcely examined issue. Not least among the merits of the essays on historical materialism included in this volume, therefore, is the fact that they provide a point of departure for this next stage in the investigation of Marx's theory of history: an inquiry into the empirical conditions for its tenability.

ANDREW LEVINE

University of Wisconsin/Madison

The Shaky Game: Einstein, Realism, and the Quantum Theory. ARTHUR FINE. Chicago: University of Chicago Press, 1986. 186 p. Cloth $25.00.

What exactly were Einstein's attitudes toward quantum theory? Do they add up to no more than a naive realism which we can no longer

0022-362X/91/8805/275–9

share? More generally, is scientific realism a defensible philosophical position? These are the questions which Arthur Fine addresses in *The Shaky Game*. The book consists of a short introductory chapter followed by eight essays written over a ten-year period. Essays 2-6 are concerned with Einstein's views on quantum theory, as they developed during the 1920s and 1930s in opposition to prevalent Copenhagen attitudes. Then, in the last three essays of the book, Fine discusses the general issue of scientific realism, and advocates a stance—he calls it the *natural ontological attitude*, or NOA—free of the rigidities he associates with realism, on the one hand, and various versions of antirealism, on the other.

As early as 1926, Fine tells us, Einstein voiced his concern about the following features of Schrödinger's theory. (1) Its equations are not relativistically invariant. (2) It does not yield classical behavior to a good approximation. (3) It leads to correlations between distant objects that violate action-by-contact principles. (4) It is an essentially statistical theory. (5) The scope of the commutation relations may not be as broad as the theory suggests. In particular we may note item (3), and that item (5) was aired *before* Heisenberg had formulated his uncertainty relations. Fine suggests that, after the Solvay conferences in the late 1920s and his discussions with Bohr, Einstein accepted that the commutation relations were indeed built into the theory; that reduced the list to four issues, concerning which Einstein was never satisfied.

In 1935, jointly with Podolsky and Rosen, Einstein published the paper, "Can Quantum Mechanical Description of Reality be Considered Complete?" (now usually known as EPR). In the same year there appeared the pair of papers by Schrödinger that introduced us to that legendary beast, Schrödinger's cat, and thereby showed how counterintuitive was the application of quantum principles to the macroscopic world. These papers are still required reading for anyone studying the foundations of quantum theory. In one of the most interesting chapters in *The Shaky Game*, Fine draws on the correspondence between Einstein and Schrödinger to reconstruct the way in which the Schrödinger-cat papers took shape, and to relate them to the EPR paper. This essay is a splendid piece of historical reconstruction.

Again, by looking at sources other than the published version of EPR (which was written by Podolsky), Fine shows that, for Einstein, the thought experiment it described had less to do with the inability of quantum mechanics to specify all "elements of reality" than with the dilemma that the theory poses: *either* a principle of separation fails to hold, *or* the theory is not complete. Thus the "principle of separation"—that the "real state" of one system cannot depend on what sort of measurement is carried out on a spatially separated system—was seen by Einstein as a key issue long before Bell made nonlocality a hot topic.

Einstein advocated a *statistical interpretation* of quantum theory, according to which the wave functions appearing in the theory describe, not single systems, but ensembles of systems. On this topic Fine makes two suggestions. Both are interesting, and both find support within Einstein's writings. There is, however, a tension between them, and my one reservation about the first part of Fine's book is that he resolves it in too facile a manner. The first suggestion is that Einstein was not interested in a hidden-variable theory, as that term is generally understood. That is, he was not interested in a supplementation of quantum mechanics *from within.* Rather, he looked for a recasting of theory that would attend to the "external" problems that quantum mechanics faces; he sought a theory which would mesh with relativity, which would have no awkward quantum/classical dichotomy built into it, and so on. One desideratum of such a theory was that in it the quantum mechanical statistics would appear as averages over ensembles, but that did not mean that the observables with which quantum theory deals were to have a special status.

Fine's other suggestion is that Einstein could escape the problems that Bell's theorem poses for ensemble interpretations by the use of *prism models*. Prism models were designed by Fine to deal with the fact that quantum mechanical probabilities cannot readily be inscribed within a classical probability space. On this approach, the probabilities associated with each measurement procedure are seen as averages over a particular subensemble of the systems tested, rather than over the whole ensemble; the whole set of quantum mechanical probabilities is then obtained by piecing together classical probability measures over different, but overlapping, subensembles.

My suspicion, however, is that the move to prism models would leave Einstein cold; it has the ring of a formal, rather than a physically motivated move. Fine acknowledges as much. Although the prism models he describes jibe with various things Einstein says—and says repeatedly—about ensemble interpretations, what they offer is effectively a hidden-variable theory, albeit one immune from Bell's challenge. As Fine says, they are clearly supplementations of the theory from within.

On the one hand, then, Einstein demands a reconstitution of physical theory, which to date remains set as a task; on the other, he speaks in ways that seem to invite the kind of supplementation that prism models provide. I see this as a genuine tension in Einstein's thought, not one that is introduced by Fine's exegesis. Indeed, Fine does a remarkable job of sifting Einstein's rather sparse and laconic utterances on these matters to reveal the two tendencies within them. Where he overstates the case on Einstein's behalf, I think, is in his resolution of the tension. When Einstein talked of an "ensemble interpretation," Fine suggests, he "chose his rhetoric cunningly" (58). In other words, those passages where he seems to be advocating, if not a prism-model approach, then one very like it, are not to be taken at face value. On the documentary evidence Fine himself presents, this seems implausible. Perhaps in this instance we should allow Einstein the luxury of being in two minds. Maybe there was an ensemble interpretation that would solve the local problems within quantum theory; ultimately, however, quantum theory would have to answer to more global demands.

In the last three essays of the book, Fine addresses the question of scientific realism. The opening sentence of essay 7, 'Realism is dead', and the title of essay 8, "And Not Anti-Realism Either," announce the dialectic of the debate. Essay 9 considers whether quantum mechanics, or any physical theory, rules out scientific realism. Fine argues, and I agree with him, that it cannot, but that the realism it permits is of a minimal kind. In what follows, I shall focus on essays 7 and 8.

The realism whose obituary appears in essay 7 is the view that the entities postulated by a successful science really exist, and that, when a branch of science consistently provides explanations of the kind we value, we should regard it as true, or approximately so; truth here is thought of as correspondence with reality. For support, realism looks to the success of scientific practice. This can only be explained, the argument runs, on the assumption that realism is true. But, as Fine points out, this argument relies on the very point at issue, to wit, whether the explanatory efficacy of a hypothesis is a warrant of its truth; for this reason, it will hardly convince the skeptic.

Fine does not, however, embrace antirealism. The varieties of antirealism he describes replace the realist's correspondence theory of truth with others: pragmatic, reductive, and so on. In contrast, Fine rejects the search for a theory of truth; we know the history, logic, and grammar of truth, he tells us, and do not need to "explain" it or to "give an account of it." Instead, the attitude he recommends (indeed, the NOA) is to say that, yes, there are electrons, but not to offer a "deeper" account of what that statement says. The NOA-er (or *knower*) simply agrees that electrons exist, but neither adds 'really' to emphasize a correspondence with some fundamental reality, nor glosses the phrase by saying, 'But, in using electron-talk, I simply mean . . .'. To ask the NOA-er what the existence claim amounts to would be to miss the point.

That is one reason why NOA is difficult to pin down. ("If I could understand it," one thinks, "I might accept it." The NOA-er, like a good Zen master, replies, "There is nothing to understand.") Another reason is that NOA has two aspects. Fine wants philosophers to reject the heuristic view that they can provide a unified account, whether realist or antirealist, of the aims, procedures, and achievements of science: the "performance" of science, he writes (148), "is not susceptible to a reading or interpretation in any global sense." We should instead let science speak for itself; our epistemic attitudes toward, for example, magnetic monopoles should be "governed by the very same standards of evidence and inference that are tolerated by science itself" (150).

We are given two prescriptions, one negative, the other positive. But the two by no means come to the same thing. In the first place, it is possible to maintain interpretive flexibility even while criticizing science from without. Fine's own book shows as much. In the second, the voices of realism and antirealism are heard, sometimes stridently, within science itself.

This brings me to my last point. Fine has given us a very good book; the historical essays are marvelously illuminating, and the arguments in the later essays are both subtle and challenging. Yet, probably because these essays reflect the movement of Fine's thought over a number of years, the two parts of the book lie somewhat uneasily together. Just as we can ask whether Socrates could find a home in the Republic, we may wonder about Einstein's fate in a world where the NOA was the one prescribed.

R. I. G. HUGHES

University of South Carolina

NEW BOOKS: TRANSLATIONS

DELEUZE, GILLES: *Expressionism in Philosophy: Spinoza.* MARTIN JOUGHIN, trans. New York: Zone, 1990. 445 p. Cloth $24.50.

HEIDEGGER, MARTIN: *Kant and the Problem of Metaphysics:* Fourth Edition, Enlarged. RICHARD TAFT, trans. Bloomington: Indiana UP, 1990. 224 p. Cloth $35.00.

IRIGARAY, LUCE: *Marine Lover of Friedrich Nietzsche.* GILLIAN C. GILL, trans. New York: Columbia, 1991. 190 p. Cloth $35.00.

KIERKEGAARD, SØREN: *Eighteen Upbuilding Discourses.* HOWARD V. HONG and EDNA H. HONG, eds. and trans. Princeton: University Press, 1990. xxii, 559 p. Cloth $59.50.

KRÄMER, HANS JOACHIM: *Plato and the Foundations of Metaphysics: A Work on the Theory of the Principles and Unwritten Doctrines of Plato with a Collection of the Fundamental Documents.* JOHN R. CATAN, ed. & trans Albany: SUNY, 1990. xxx, 311 p. Cloth $49.50, paper $16.95.

LEVINAS, EMMANUEL: *Nine Talmudic Readings.* ANNETTE ARONOWICZ, trans. Bloomington: Indiana UP, 1990. xxxix, 197 p. Cloth $29.95.

NEGRI, ANTONIO: *The Savage Anomaly: The Power of Spinoza's Metaphysics and Politics.* MICHAEL HARDT, trans. Minneapolis: Minnesota UP, 1991. xxiii, 277 p. Cloth $44.95, paper $18.95.

RICKEN, FRIEDO: *Philosophy of the Ancients.* ERIC WATKINS, trans. Notre Dame: University Press, 1991. x, 280 p. Cloth $24.95, paper $12.95.

VICO, GIAMBATTISTA: *On the Study Methods of Our Time.* ELIO GIANTURCO, trans. With *The Academies and the Relation Between Philosophy and Eloquence,* DONALD PHILLIP VERENE, trans. Ithaca: Cornell, 1990. xlix, 90 p. Cloth $24.95, paper $6.95.

THE JOURNAL OF PHILOSOPHY
is pleased to announce the publication of its
TWENTY-FIVE YEAR INDEX, 1964–1988
VOLUME LXI TO VOLUME LXXXV
AUTHORS, SUBJECTS, & REVIEWS
Supplement to its
FIFTY-YEAR INDEX, 1904–1953
TEN-YEAR SUPPLEMENT, 1954–1963

It is available for the rest of 1991 at special prices
25-year (1964–1988) $25 00 cloth/$10 00 paper
also available
50-year (1904–1953) $12 00 cloth 10-year (1954–1963) $3 50 paper
Shipping $5 00 (first)/$2.50 (each addl)

Please send a check or money order to

Journal of Philosophy
704 Philosophy Hall
Columbia University
New York, New York 10027

Prices are in effect through December 31, 1991. Orders and payment must be received by then

THE ANCIENT QUARREL BETWEEN POETRY AND PHILOSOPHY

Thomas Gould uses both historical and theoretical approaches to explore tragedy and its power to gratify readers and audiences. He takes as his starting point Plato's moral and psychological objections to tragedy, and the conflict he recognized between "poetry"—the exploitation of our yearning to see ourselves as victims—and "philosophy"—the insistence that all good people are happy. Plato's objections to tragedy are shown to be an essential feature of Socratic rationalism and to constitute a formidable challenge even today.

"Gould's book is a significant contribution to the study of tragedy, both ancient and modern, to our understanding of Plato and Aristotle, and to the application of psychoanalysis to literature. His argument is powerful. I felt as I read that many things were falling into place for me."
—*William G. Thalmann,*
University of Southern California

Cloth: $39.50 ISBN 0-691-07375-9

New in paperback

For Self-Examination/ Judge for Yourself!

Søren Kierkegaard

Edited and translated, with Introduction and notes, by Howard V. Hong and Edna H. Hong

For Self-Examination and its companion piece *Judge for Yourself!*, now available in a definitive paperback edition, are the culmination of Søren Kierkegaard's "second authorship," which followed his *Concluding Unscientific Postscript*. Among the simplest and most readily comprehended of Kierkegaard's books, the two works are part of the signed direct communications, as distinguished from his earlier pseudonymous writings. They contain the well-known passages on Socrates' defense speech, how to read, the lover's letter, the royal coachman and the carriage team, and the painter's relation to his painting.

Now in paper: $12.95 ISBN 0-691-02068-3

Princeton University Press
41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 258-4900
ORDERS: 800-PRS-ISBN (777-4726) • OR FROM YOUR LOCAL BOOKSTORE

Cambridge University Press

The Philosophy of F.P. Ramsey

Nils-Eric Sahlin

"Sahlin has touched on all the bases not only providing us with an accurate exposition of Ramsey's ideas but with illuminating discussion of their significance."

— *Isaac Levi, Columbia University*

38543-1 Hardcover $39.50

Now in paperback...

The Empire of Chance

How Probability Changed Science and Everyday Life

Gerd Gigerenzer, et al.

"... will be useful to statisticians, philosophers, scientists and other historians of science who want to understand the roots of the probability-based statistical methods we use so widely today... *The Empire of Chance* is a valuable book."

— *Science*

Ideas in Context

Copublished with the support of the Exxon Education Foundation

39838-X Paper $16.95

Sensations

A Defense of Type Materialism

Christopher S. Hill

"Part I does an excellent job of laying out the problematic for the book [and] sets an extremely high standard of clarity that is remarkably sustained throughout the rest of the book. Part II contains both the best criticism of functionalism and sensations of which I am aware and the best defense of type materialism of which I am aware." — *Brian McLaughlin, Rutgers University*

Cambridge Studies in Philosophy

39423-6 Hardcover $44.50
39737-5 Paper $15.95

Vico

A Study of the "New Science"

Second Edition

Leon Pompa

A new chapter that analyzes Vico's conception of the development of law has been added to this edition of a unique work devoted almost exclusively to an interpretation of the New Science.

38217-3 Hardcover $44.50
38871-6 Paper $12.95

Fichte's Theory of Subjectivity

Frederick Neuhouser

"[A] very substantial piece of scholarship which analyzes a number of important historical and systematic issues with great clarity and perception. The presentation and treatment of the basic historical and philosophical issues is magisterial."

— *Raymond Geuss, Columbia University*

37433-2 Hardcover $39.50
39938-6 Paper $12.95

Leibniz's Philosophy of Logic and Language

Second Edition

Hide Ishiguro

"All serious students of Leibniz's philosophy will have to read her book."

— *Times Literary Supplement*

This edition of an important introduction to Leibniz's philosophy of logic and language, first published in 1972, questions traditional interpretations and reveals how his thought is related to current logic theory interests.

37428-6 Hardcover $49.50
37781-1 Paper $16.95

The Chain of Change

A Study of Aristotle's *Physics* VII

Robert Wardy

The Chain of Change is the first full-scale philosophical commentary devoted to Aristotle's *Physics* VII, in which Aristotle argues for the existence of a first, unmoved cosmic mover. This study systematically considers the major issues of the book, and argues for the fundamental importance of *Physics* VII in our understanding of Aristotelian cosmology and natural science.

Cambridge Classical Studies

37327-1 Hardcover $59.50

Available in bookstores or write:
CAMBRIDGE
UNIVERSITY PRESS
40 West 20th Street,
New York, NY 10011.
Call toll-free 800-872-7423.
MasterCard/VISA accepted.
Prices subject to change.

KLUWER ACADEMIC PUBLISHERS

including the D. Reidel Publishing Company
and Martinus Nijhoff Publishers imprints

Publishers of leading journals in:

Philosophy of Science
Logic
Phenomenology
History of Philosophy
Epistemology
Philosophy of Medicine
Philosophy of Law
etc.

Our journals include:

Studies in Philosophy and Education. Editor: David Ericson
Human Studies. Editor: George Psathas
Journal of the History of Biology. Editor: Everett Mendelsohn
Man and World. Editors: John M. Anderson, Joseph Kockelmans and Calvin Schrag
Philosophical Studies. Editor-in-chief: Keith Lehrer
Synthese. Editor-in-chief: Jaakko Hintikka
Theory and Decision. Editor-in-chief: W. Leinfellner
Argumentation. Editor-in-chief: Michel Meyer
Biology & Philosophy. Editor: Michael Ruse
Husserl Studies. Editors: J.N. Mohanty and Karl Schuhmann
Law and Philosophy. Editor: Alan Mabe
The Journal of Medicine and Philosophy. Editors: Tristam Engelhardt and Stuart Spicker
The Journal of Philosophical Logic. Editors: J.Michael Dunn and Terrence Parsons.
The Journal of Value Inquiry. Editor: Robert Ginsberg
Theory and Society. Editor: Janet Gouldner
The Journal of Business Ethics. Editor: Alex Michalos
International Journal for Philosophy of Religion. Editor: Eugene Long
Minds and Machines. Editor: James Fetzer
Linguistics and Philosophy. Editor: David Dowty

For more information on our publishing programs, contact:
Kluwer Academic Publishers
101 Philip Drive · Norwell, MA 02061 · (617) 871-6600

Classics in Philosophy

GREEK PHILOSOPHY
2nd Edition
Reginald E. Allen
ISBN: 0-02-900660-0 $12.95 paper

POLITICAL IDEAS OF ST. THOMAS AQUINAS
St. Thomas Aquinas
ISBN 0-02-840380-0 $10.95 paper

EIGHTEENTH CENTURY PHILOSOPHY
Lewis W. Beck, Editor
ISBN 0-02-902100-6 $12.95 paper

A MODERN INTRODUCTION TO PHILOSOPHY, 3rd Edition
Paul Edwards & Arthur Pap, Editors
ISBN 0-02-909200-0 $27.95

A SHORT HISTORY OF CHINESE PHILOSOPHY
Fung Yu-Lan
ISBN: 0-02-910980-9 $14.95 paper

NINETEENTH-CENTURY PHILOSOPHY
Patrick Gardiner, Editor
ISBN: 0-02-911220-6 $12.95 paper

INTRODUCTORY READINGS IN AESTHETICS
John Hospers
ISBN 0-02-915260-7 $14.95 paper

MORAL AND POLITICAL PHILOSOPHY
David Hume
ISBN 0-02-846170-3 $12.95 paper

DIALOGUES CONCERNING NATURAL RELIGION
David Hume
ISBN 0-02-846180-0 $10.95 paper

ESSAYS IN PRAGMATISM
William James
ISBN. 0-02-847140-7 $10.95 paper

CRITIQUE OF JUDGMENT
Immanuel Kant
ISBN 0-02-847500-3 $12.95 paper

JOHN STUART MILL'S PHILOSOPHY OF SCIENTIFIC METHOD
John Stuart Mill
ISBN 0-02-849250-1 $12.95 paper

PHILOSOPHY AND TECHNOLOGY
Carl Mitcham and Robert Mackey, Editors
ISBN 0-02-921430-0 $14.95 paper

THE SPIRIT OF THE LAWS
Baron de Montesquieu
ISBN 0-02-849270-6 $13.95 paper

NEWTON'S PHILOSOPHY OF NATURE
Isaac Newton
ISBN 0-02-849700-7 $12.95 paper

A CRITICAL HISTORY OF WESTERN PHILOSOPHY
D.J. O'Connor
ISBN 0-02-923840-4 $14.95 paper

PHILOSOPHY OF THE SIXTEENTH AND SEVENTEENTH CENTURIES
Richard H. Popkin
ISBN 0-02-925490-6 $12.95 paper

GREEK AND ROMAN PHILOSOPHY AFTER ARISTOTLE
Jason L. Saunders
ISBN 0-02-927730-2 $14.95 paper

ETHICS
Benedict de Spinoza
ISBN 0-02-852650-3 $10.95 paper

TWENTIETH CENTURY PHILOSOPHY
The Analytic Tradition
Morris Weitz
ISBN: 0-02-934990-7 $12.95 paper

MEDIEVAL PHILOSOPHY
John F. Wippel & Allan B. Wolter
ISBN 0-02-935650-4 $12.95 paper

Attention College Professors: Examination copies will be sent on a 30-day approval basis only. Simply place your request on university stationery, indicating book requested, course title, projected enrollment, and date your next decision is due. When ordering examination copies of paperbacks, enclose a check for $2.00 per copy to cover postage and handling. Mail to: The Free Press, Attn. Lisa Weinberg, 866 Third Avenue, NY, NY 10022

Fp **THE FREE PRESS**
A Division of Macmillan, Inc.
866 Third Avenue, New York, New York 10022

INTRODUCTION TO METAPHYSICS

The Fundamental Questions

edited by Andrew B. Schoedinger

Comprehensive in its selection of readings, this volume is the most complete collection of seminal essays on metaphysics available today. Divided into five sections, the essays address the issues that make metaphysics such a challenging and rewarding subject: from the concept of universals to the thorny problem of causation, from understanding what is meant by personal identity to unraveling the issues surrounding free will to the philosophical challenges presented by artificial intelligence. Includes selections by Aristotle, R. Bambrough, George Berkeley, Joseph Butler, Rudolf Carnap, Roderick Chisholm, Arthur Danto, Donald Davidson, Rene Descartes, C. J. Ducasse, David Hume, John Locke, Alastair MacIntyre, A. I. Melden, John Stuart Mill, Anthony Quinton, D F. Pears, Hilary Putnam, Bertrand Russell, Sydney Shoemaker, P. F. Strawson, Richard Taylor, and many others.

375 pages • Paper $21.95

THE QUESTION OF HUMANISM

Challenges and Possibilities

edited by David Goicoechea, John Luik, and Tim Madigan

What does the term "humanism" mean today? What is the nature of the views it represents? How can it meet the challenges of contemporary life? This absorbing collection of original essays examines the abundant variety of historical and contemporary humanist philosophies. Emphasis is placed on the works of such philosophers as Thomas Aquinas, Immanuel Kant, Søren Kierkegaard, Karl Marx, Friedrich Nietzsche, Martin Heidegger, and Michel Foucault. Included are essays by Cecil Abrahams, Zygmunt Adamczewski, Samuel Ajzenstat, Martin Andic, Allan Booth, Richard S. G. Brown, Michael Cardy, Kenneth Dorter, Richard Francis, David Goicoechea, Danny Goldstick, Calvin Hayes, Martha Hewitt, Monica Hornyansky, Paul Kurtz, James Lawler, John Luik, Tim Madigan, Robert McLaughlin, Graeme Nicholson, Zaid Orudjev, Robert Perkins, Charles Scott, and Edward A. Synan.

342 pages • Cloth $29.95

PHILOSOPHICAL ESSAYS IN PRAGMATIC NATURALISM

by Paul Kurtz

This volume highlights the thrust and focus of the two dominant schools of 20th century American philosophy. Special emphasis is given to empirical metaphysics, a modified naturalistic ethic, and a critique of phenomenology and existentialism.

"Seminal ideas on bridging the rift between holism and reductionism." —Publishers Weekly

246 pages • Cloth $19.95

PHILOSOPHICAL EXPLORATIONS

Freedom, God, and Goodness

by Steven M. Cahn

Presents a fascinating discussion of three of the most talked about topics in Western thought: freedom of the will, the existence of God, and the reasonableness of value judgments.

200 pages • Paper $14.95

At better bookstores, or order directly from

PROMETHEUS BOOKS

Call Toll Free (800) 421-0351 (24 hours)

59 John Glenn Drive •Amherst, NY 14228-2197

Add $3.00 p&h for first book, $1.50 for each additional book ($9.00 maximum). NY State residents add sales tax. #40053

Crossings

Nietzsche and the Space of Tragedy

John Sallis

Boldly contesting recent scholarship, Sallis argues that *The Birth of Tragedy* is a rethinking of art at the limit of metaphysics. His close reading focuses on the complexity of the Apollinian/Dionysian dyad and on the crossing of these basic art impulses in tragedy.

Paper $11.95 192 pages
Library cloth edition $24.95

The Rhetoric of Morality and Philosophy

Plato's Gorgias *and* Phaedrus

Seth Benardete

Distinguished classicist Seth Benardete here interprets and, for the first time, pairs two important Platonic dialogues, the *Gorgias* and the *Phaedrus*. In linking these dialogues, he places Socrates' notion of rhetoric in a new light and illuminates the way in which Plato gives morality and eros a place in the human soul.

Cloth $29.95 215 pages

Now in Paper

Of Spirit

Heidegger and the Question

Jacques Derrida

Translated by Geoffrey Bennington and Rachel Bowlby

"Will a more important book on Heidegger appear in our time? No, not unless Derrida continues to think and write in his spirit. . . . Let there be no mistake: this is not merely a brilliant book on Heidegger, it is thinking in the grand style."—David Farrell Krell, *Research in Phenomenology*

Paper $9.95 148 pages

٣/٩/٩١

THE UNIVERSITY OF CHICAGO PRESS

5801 South Ellis Avenue Chicago, IL 60637

THE JOURNAL OF PHILOSOPHY

philosophy

VOLUME LXXXVIII, NUMBER 6
JUNE 1991

page
281 *Contractualism, Moral Motivation, and Practical Reason* Samuel Freeman

COMMENTS AND CRITICISM
304 *The Causal History of Computational Activity: Maudlin and Olympia* Eric Barnes

BOOK REVIEWS
317 Stuart Hampshire: *Innocence and Experience* David Luban
324 Eva Kittay: *Metaphor: Its Cognitive Force and Linguistic Structure* Peter Ludlow
331 Michael Detlefsen: *Hilbert's Program: An Essay on Mathematical Instrumentalism* Mark Steiner

CENTRAL LIBRARY

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secrėtary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1991 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER. Second class postage paid at New York, NY, and additional mailing offices Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NO. 6, JUNE 1991

CONTRACTUALISM, MORAL MOTIVATION, AND PRACTICAL REASON*

John Rawls suggests in *A Theory of Justice*[1] that the contractarian idea might be extended to (1) explicate the concept of right, and (2) argue for "the choice of more or less an entire ethical system . . . including principles for all the virtues and not only for justice" (*ibid.*, pp. 17, 111; cf. pp. 184, 584–5). T. M. Scanlon[2] has outlined an account that in many ways responds to Rawls's suggestions. Contractualism, he contends, offers (1) a "foundational account" of the subject matter of morality, and (2) a general moral view. Scanlon's account is incomplete. But it presents a conception of morality many will find compelling as an alternative to both noncognitivism and utilitarianism. Especially appealing is the contractualist account of moral motivation, since it provides a way to understand obscure but central aspects of Kantian views: the claims that moral reasons are of a special kind, and that moral motives have a basis in practical reason. My aim is to develop Scanlon's account to show how it exhibits these features. Contractualism's primary significance lies in the way it relates motivation to practical reason. This is the subject of sections II–III. To provide necessary background, I begin with a discussion of Scanlon's account. He presents contractualism mainly by contrast with the philosophical bases of utilitarianism. But to appreciate how distinctive contractualism is requires its comparison with other foundational and normative views.

---

* I am indebted to T. M. Scanlon, Burton Dreben, and Jay Wallace for helpful comments on drafts of this paper. My greatest debt is to John Rawls, especially to his unpublished lectures on Kant's ethics. A research grant from the University of Pennsylvania enabled me to write this article.

[1] Cambridge: Harvard, 1971. Hereafter cited as TJ.

[2] "Contractualism and Utilitarianism," in *Utilitarianism and Beyond*, Amartya Sen and Bernard Williams, eds. (New York: Cambridge, 1982), pp 103–28. Hereafter cited as CU

0022-362X/91/8806/281–303

I. CONTRACTUALISM AND THE FOUNDATION OF MORALITY

Characteristic of all social-contract views are the ideas that social rules and institutions are to be freely acceptable to all persons bound by them, and for their mutual benefit. These intuitive notions give rise to the idea that moral requirements are grounded in a general agreement. There are different ways to arrive at and describe free acceptability and mutual benefit. In Rawls, they follow from his specification of a conception of social cooperation and a democratic conception of citizens as free and equal moral persons. David Gauthier interprets these ideas starting from an account of rationality as utility maximization, and cooperation-free interaction in a state of nature. As much as these contract views differ, both are directed toward developing a substantive account of political morality. Scanlon's perspective on these ideas (free acceptability, mutual benefit, and general agreement) is different. He sees them, in the first instance, as part of a philosophical account of "the nature of morality" (CU, 110).

Any normative conception of ethics must provide, Scanlon contends, a response to three fundamental philosophical questions:

(1) A foundational question: What is the subject matter of morality, and the grounds of truth, if any, for moral judgments?
(2) An epistemological question: How can we come to know these grounds, i.e., ascertain the kinds of facts that moral truths are about?
(3) A motivational question: Why should anyone care about morality, and be moved to act on its demands (CU, 104–6)?

Utilitarianism derives much of its force, Scanlon says, from the fact that it relies on philosophical premises that give relatively clear-cut answers to these questions (CU, 108). Scanlon calls this view "philosophical utilitarianism," to distinguish it from "normative utilitarian" conceptions, all of which hold the substantive moral thesis that it is right to maximize the sum of well-being. Philosophical utilitarianism is basically a claim about the subject matter of morality; it is "the thesis that the only fundamental moral facts are facts about individual well-being" (CU, 108).

This thesis should not be confused with the idea that morality must advance human interests, their well-being, or their good. These claims are endorsed by a wide range of views—not just utilitarianism and perfectionism, but also Rawls's and Scanlon's Kantian views, and by Kant himself. Some may deny morality need advance the human good (e.g., Ross and Prichard are often read this way). But this is not the real issue between teleological (or consequentialist) and deontological theories. Nor is it the question whether consequences of actions matter in assessments of what it is right to do. (As Rawls says, any view that denies that "would simply be irrational, crazy" (TJ, 30).) Instead, underlying the rift between teleological and deontological (or at least Kantian) views are different conceptions of practical reasoning. They approach differently the question: How are considerations of individuals' good to enter into moral reasoning, and how are we to assess and promote their well-being? Philosophical utilitarianism implies a particular response here: If (1) facts about well-being are the *only* fundamental moral facts, and we assume that (2) these facts can be defined independently of moral concepts and principles, then we are led to the idea that (3) morality has no intrinsic value of its own, but is simply an instrument (one we would be better off without if things were different) for advancing the human good. This consequence, when joined with the normative claim that (4) no one's well-being counts for more than anyone else's, leads to the primary thesis of normative utilitarianism that (5) in appraisals of what is right to do ultimately we should be guided by what maximizes the sum total of individual well-being (cf. CU, 109-10).

Scanlon especially wants to reject the idea that all moral questions are ultimately reducible to decisions about what conduct and institutions maximize an aggregate. Something very different is involved in right conduct, just institutions, and in moral reasoning even at the most abstract levels. What this is, the underlying idea of Scanlon's contractualism, is best approached by focusing on the role of moral requirements in social interaction.

Every social group has its system of norms, certain rules and practices which regulate the conduct of members of the group and which are seen as essential to continued social cooperation. If we concentrate solely on the role of these norms in guiding conduct and regulating interaction, we are led to the idea that the purpose of moral requirements is to maintain and promote individuals' well-being. But exclusive attention to the action-guiding aspects of moral norms can lead to disregard of an equally significant function: they also have a place in the public life of the group in that they provide a common basis for argument and agreement among members regarding one another's conduct and expectations. In the course of justification within any social group, moral norms serve as the primary source of *public reasons,* the considerations we appeal to in order to justify and assess critically one another's actions on the basis of generally accepted criteria. As such, moral norms provide the basic framework for the assessment of individual and institutional conduct, its justification, criticism, and praise and blame.[3]

We might look upon this public role of systems of moral norms[4] in providing a basis for public justification and agreement as derivative of its regulative role. That people are able to justify their conduct is normally necessary for social stability, a condition of individual well-being on any account of morality. An example of a purely instrumental approach to the public role of moral rules is the tendency of some utilitarians[5] to depict moral praise, blame, and criticism as actions just like any other, to be assessed wholly in terms of their usefulness to aggregate well-being. Here the role of rules in public justification collapses; what we publicly *say* to one another to assess conduct is to be assessed (presumably nonpublicly) in terms of the conduciveness of our pronouncements to overall well-being.

In stark contrast, contractualism assigns independent significance to morality's role in providing a framework for public justification, and seeks to account for the regulative purposes of moral norms through it. In answer to the question, "To what end are we to regulate our conduct?" the response is not "well-being," but "So that we may be able to justify our conduct to one another by considerations that everyone can reasonably accept."

Scanlon's most succinct characterization of contractualism is in these terms, in "an example" of "the kind of property which moral wrongness is":

> An act is wrong if its performance under the circumstances would be disallowed by any system of rules for the general regulation of behaviour which no one could reasonably reject as a basis for informed, unforced general agreement (CU, 110).[6]

---

[3] This says nothing about the validity of conventionally accepted moral norms, or their bases in self-deception or ideology. It simply provides a different approach to morality than utilitarianism or Hobbesian views. For further discussion of public reasons, see my "Reason and Agreement in Social Contract Views," *Philosophy and Public Affairs*, XIX, 2 (Spring 1990)· 122–57.

[4] On the related idea of the social role of moral principles, see Rawls, "Kantian Constructivism in Moral Theory," The Dewey Lectures 1980, this JOURNAL, LXXVII, 9 (September 1980): 515–72, esp. p. 517.

[5] Henry Sidgwick, *Methods of Ethics*, 7th ed. (London: MacMillan, 1907), pp. 428–9; and J. J. C Smart, "Freewill, Praise and Blame," *Mind*, LXX (1961): 291–306.

[6] More recently, Scanlon has characterized "a contractualist account of morality" as "a theory according to which an act is right if it would be required or allowed by principles which no one, suitably motivated, could reasonably reject as a basis for informed, unforced general agreement." "The Significance of Choice," *The Tanner Lectures on Human Value* (Salt Lake City: Utah UP, 1987), pp. 149–216, esp. p. 151. (Hereafter SC.)

Central to this definition is the idea that individuals are answerable to one another for their conduct, the claims they make, and the expectations they have regarding others' conduct and claims.[7] This is part of what it means to be reasonable. Answerability implies more than being held accountable (responsible) for one's acts. It means that we (or someone in our stead) be able publicly to justify our conduct, aims, and expectations on terms others could freely accept. This gives rise to the contractualist idea that the norms we rely on to justify as well as to regulate our actions and expectations are the product of an ideal agreement.

More needs to be said about the idea of reasonableness. But, first, consider the objection that there is nothing in Scanlon's definition, being reasonable, or even seeking a basis for public agreement, that would warrant representing morality as the *product* of a general agreement. Standards of reasonable acceptability and public agreement are *better* met by appeal to a prior realm of moral facts independent of us, which are evident to the light of informed reason (or to moral sense). If such facts exist, then rational and informed individuals would all, trivially, agree on them. They agree because all have an undistorted view of this independent moral realm. If so, there is no need to invoke a separate agreement as the *basis* for public agreement and what is reasonable; agreement would play no fundamental role in defining the subject matter of morality or the content of moral principles.

Scanlon says, "the appeal of contractualism . . . rests in part on a qualified skepticism" (CU, 118), by which I assume he means a rejection of an independent realm of moral facts and a special faculty of reasoning to ascertain them. This does not imply skepticism about the truth value of moral claims. Indeed, Scanlon's real point is that, if we see morality in terms of a hypothetical agreement among idealized individuals, then there is *no need* to invoke an antecedent realm of moral facts to provide a basis for moral truth and cognition. To clarify: imagine a community of free and informed agents, each of whom is sincere, conscientious, and motivated by a desire publicly to justify his actions, ends, and expectations to everyone else. Morality is the set of principles and rules which these persons could all affirm and to which they could jointly commit themselves, knowing that these terms (a) are to regulate their actions and expectations, and (b) are to serve as the ultimate public standards for justification, criticism, and the adjudication of conflicting claims.[8]

---

[7] Annette Baier suggested to me the centrality of answerability here.

[8] Although Scanlon does not explicitly define contractualism this way, I believe it incorporates the central features of his view—cf CU, 111. The parallel with

As an account of morality's subject matter, Scanlon's view implies, not that our moral judgments are true when they match the judgments on which rational, informed individuals who have a privileged view of a prior moral order can agree. Instead, moral truth is defined by reference to the principles to which free and informed persons could reasonably agree under ideal conditions. There *are* no moral facts or properties prior to this hypothetical agreement. Significantly, this account provides a definition of moral truth and moral objectivity which does not require the ontological and epistemological commitments of rational intuitionism or its analogues. Contractualism thereby avoids the argument of J. L. Mackie's error theory (that there are no moral facts "in the world" for our moral judgments to be true of) (cf. CU, 108). We might then see contractualism as frustrating subjectivists' claims to be the spokesmen for the natural sciences. While it provides a conception of moral truth different from any that is appropriate for scientific claims, nothing in the account conflicts with descriptions of nature offered by natural science.

But contractualism must do more than this. Mackie's objection also might be avoided by a moral-sense view that is not committed to a realm of moral facts independent of human sensibilities. The main example here is ideal-observer theories, which define moral truth by the emotional responses of an ideally situated being with full knowledge and a perspicious view of the universe.[9] These accounts share in common with rational intuitionism the assumption that conditions of *observation* of some kind are fundamental to making correct moral judgments. This is the broader issue addressed by contractualism's focus on ideal judgment and agreement. It is not clear what the relevance of appeals to observation—no matter how ideal its conditions—are to grounding the truth of moral judgments.[10] In the context of moral justification, such appeals carry

---

Kant's third formulation of the categorical imperative, in terms of a "Kingdom of Ends," deserves separate discussion. Cf. TJ, 264.

[9] See Roderick Firth, "Ethical Absolutism and the Ideal Observer," *Philosophy and Phenomenological Research*, XII (March 1952). Others have lately relied on a similar idea for accounts of moral truth and objectivity. Gilbert Harman, "Moral Explanations of Natural Facts—Can Moral Claims be Tested against Moral Reality," *Southern Journal of Philosophy*, XXIV, Supplement (1986): 57–68; Peter Railton, "Moral Realism," *Philosophical Review*, XCV, 2 (April 1986): 190ff. David Lewis and Michael Smith, "Dispositional Theories of Value," in *Proceedings of the Aristotelian Society, Supplement*, 1989.

[10] Rawls raises this problem early on, in "Justice as Fairness," *Philosophical Review*, LXVII (April 1958): 164–94, esp. p. 194n; and it informs the account of objectivity in his "Kantian Constructivism in Moral Theory," pp. 554–72.

little argumentative force. Early in his essay, Scanlon says of both morals and mathematics that we need an epistemology appropriate to the subject matter of each, which explains "how it is possible to discover the facts about this subject matter *through something like the means we seem to use*" (CU, 104, emphasis added). Appeals to observation and a realm of moral facts prior to and independent of *reasoning* are not among the means we actually use in moral justification. In this context, observational appeals provide "reasons" of the wrong kind. To say there are moral qualities "out there in the world" or implicit in my perception of situations, which I detect by intuition or moral sense, does not provide a public justification of moral judgments, but is a way to cut short the process of justification. In moral argument, we seek to discover, not statements describing antecedent facts, but principled justifications of our moral convictions. This is the work that is done by appeal to ideal conditions for judgment and agreement.

This does not mean the notions of "moral fact," "intuition," etc., have no place in moral thought and argument. But the role they do have is different from the role of appeals to facts and experience in theoretical contexts. Appeals to moral facts, or to the observations of moral experience, are not appeals to brute data. They call upon the well-considered and firmly held moral convictions we share. And we share these convictions for a *reason*. The reasons that inform our intuitions, in moral contexts, are ultimately provided by principles.[11] Contractualism accounts for these principles by appealing to certain features of our practices of public justification. It explains morality by idealizing the ways of reasoning and justification we normally use in moral deliberation and argument. On this basis—and not by analogy with theoretical reasoning and its ideal conditions—contractualism constructs an "interpersonal standpoint" (CU, 117) that provides an objective point of view from which to ground claims about the validity of our moral judgments and conventions, and to assess arguments for competing moral principles. Our moral judgments are true when they accord with principles everyone could judge most reasonable from this perspective. Unlike rational intuitionism and ideal-observer theories, there is no assumption that moral facts or properties exist prior to and independently of *judgment* from this objective perspective. Rather, judgment from this standpoint specifies the realm of moral facts. Objectivity is, as it were, prior to the universe of moral "objects." The principles that ideally situated individuals could agree to define what the properties of rightness and wrongness are, and determine what natural and institutional facts are to be taken by us as morally relevant in our deliberations.[12]

---

[11] For example, what makes total domination of a person, or killing for enjoyment, morally relevant facts is that they violate principles we share. We say why these facts are morally relevant by appealing to these principles

The controversy between moral realism and moral antirealism is a complicated subject. My intention is but to assign contractualism a place in this debate by indicating its distinct approach to these issues. Contractualism defines moral truth and objectivity, not in terms of antecedent properties (natural or non-natural), or ideal observational conditions, but by idealized conditions for judgment and agreement that model the nature and structure of our practical reasoning. (How it does this will be discussed in II below.)

Turn now to a third comparison, to clarify the role of the idea of *reasonableness*. One can see how, given contractualism's emphasis on public justification, this idea acts as a counterpoise to the utilitarian thesis that morality involves maximizing aggregate well-being. But reasonableness does more than alleviate the additive aspects of utilitarian and consequentialist views. Contrast contractualism with Gauthier's Hobbesian account. Free acceptability and mutual benefit are elements of any contract view. How these ideas are related marks a significant difference between contract theories. Scanlon sees them as distinct and independent notions; Gauthier does not. He represents rational beings as agreeing to certain moral constraints on the direct pursuit of their ends, solely for the purpose of more effectively maximizing their individual well-being. Free acceptance of these terms is one means among others to achieve this benefit. In this way, Gauthier's[13] contract theory is more akin to philosophical utilitarianism than to contractualism; it holds that facts about individual well-being are the only fundamental moral facts. This resemblance is reflected in the parallel ways Gauthier and utilitarians treat individuals' desires. Both take desires as given, and as having value on their own terms, without regard to their origin or what they are desires for. Moreover, both define rationality in maximizing terms. The difference is that Gauthier refuses to aggregate satisfactions. There lies the significance of his appeal to a general agreement. As a bargain, the purpose of Gauthier's agreement is not to provide public answerability for the claims individuals make in the name of their well-being. Self-interested utility maximizers have no concern for that, or for the adverse effects their claims have on the good of others.[14] They have no desire to be reasonable.

---

[12] Here I interpret Scanlon, CU, 118. Rawls makes a similar point in "Kantian Constructivism," pp. 564–5, 568–9; see also, TJ, 45, 516f.

[13] "Moral theory is essentially the theory of optimizing constraints on utility-maximization"—*Morals by Agreement* (New York: Oxford, 1986), p. 78.

For Scanlon, general agreement is not instrumental to well-being; "it is, in a more fundamental sense, what morality is about" (CU, 128). His agreement is not then represented as a bargain or compromise among essentially conflicting interests; the kind of agreement morality involves is different. Moreover, individual well-being is morally significant, "not because it is intrinsically valuable . . . but simply because an individual could reasonably reject" in moral argument to general agreement, principles "that gave his well-being no weight" (CU, 119). Contractualism's denial of the intrinsic value of well-being means desires are not taken as given in moral evaluation, and assigned importance simply on the basis of their strength. Contractualism assumes an account of individuals' legitimate interests—desires that are (in Kant's terms) *worthy* of satisfaction—which is defined by reference to the codeliberations of Scanlon's parties and the moral principles to which they agree. I shall discuss these features in the course of elaborating on Scanlon's account of moral motivation and its relation to practical reasoning.

II. CONTRACTUALISM AND MOTIVATION

For any philosophical account of morality to be convincing, it has to connect awareness of moral requirements with action. An account of the nature of morality should show why people are regularly motivated to do what they believe to be right. And a normative moral theory needs to show why people should be moved to act on its substantive requirements. Both problems raise questions as to "why anyone should care about" morality (CU, 105), as defined by an account of its subject matter, or by the principles of a moral theory.

To lay the background for Scanlon's account of moral motivation, we need distinguish two different ways to think of reasons as instrumental, relied on to support two accounts of rationality. We often speak of the reasons a person has by reference to his primary desires. On one philosophical account, which I shall call "Humean" or "desire-based," a necessary condition of *A*'s having a reason to do some act is that the act advance the satisfaction of a present desire *A* has. Most accounts of this sort allow that one may have reasons of which he is unaware. Moreover, a good deal of idealization may be required to say what one's current desires, and thus reasons, are

---

[14] Gauthier's parties' lack of concern for the reasonableness of their claims comes out in their initial bargaining demands in the social contract; each demands the maximum, the entire cooperative surplus—*ibid.*, pp. 141–50

(e.g., what one would want consistently and after full consideration). But the basic assumption of Humean views is that, for each person *A*, there is a set of primary desires he has at any time *t*, which are given to him by his nature and circumstances, and which provide the necessary basis for his having any reasons for acting. Gauthier, Gilbert Harman, Bernard Williams, and others hold this view.

We also often think of reasons less subjectively. For example, some people (e.g., children) have reason to provide for their health, education, or future, though they have no concern for them now. This presumes people have interests independently of present desire. Based on this, "interest-based accounts" say reasons are at least partially defined by certain (objective) interests. This presupposes a notion of persons' rational good. Reasons are still instrumental, so long as they are definable by reference to what promotes one's antecedent interests. Philippa Foot has held a version of this view. She contends that prudential considerations give agents reasons, without respect to their current desires.

On the basis of both conceptions, it has been argued (by Foot, Williams, Harman, Gauthier, and others) that the kinds of reasons moral considerations provide are instrumental. One can have reason to act morally only if it satisfies a primary desire with moral content (e.g., a desire to be just, fair, honest) or at least advances some other desire or nonmoral interest. If so, the kinds of obligations, hence reasons, to which moral considerations give rise are, at best, "hypothetical." No sense can be made of Kant's idea that moral reasons are "categorical" or of a special kind.[15]

Now, Scanlon seems to deny that either of these accounts of reasons is adequate. But he does not deny that we can have desire- and interest-based reasons (CU, 105). The two motivational questions he raises are framed just in these terms. The first question is: What is "the nature of the reasons that morality does provide, at least to those who are *concerned* with it" (CU, 105-6)? The second asks whether having moral motives "is compatible with a person's good or whether it is, as Nietzsche argued, a psychological disaster for the person who has it" (CU, 106). My concern here is the first question.

What is Scanlon getting at? He is not trying to show the amoralist, the person with no moral concerns, that he has reason to do what is right whatever his desires may be. The Hobbist addresses this question. The problem Scanlon raises is different: How are we to account for the *fact* that moral considerations have such a grip on us?

---

[15] See Philippa Foot, *Virtues and Vices* (Berkeley: California UP, 1978), chs. 10–2.

Scanlon takes it for granted that moral concerns move the great majority of people; moreover, "they strike those who are moved by them as reasons of a special stringency and inescapability" (CU, 106). Given this, his questions are: (1) What desire do we *already* have which leads us, almost universally, to take so seriously our awareness of moral matters (as opposed to other facts)? (2) What is there about the object of this desire which warrants such special concern?

One problem with rational intuitionism is that it provides no answer to (2). Intuitionism holds that the recognition of certain moral facts (by one capable of intuiting them) is sufficient to give rise to a desire to do what is right *simply because* it is right. Moral motivation is defined by reference to a special cause, the intuitive grasp of non-natural properties or first principles.[16] Hume's argument against this view is based in the features of his account of human nature (reason is "inert" and so cannot give rise to any desire). One might contest that account (arguing that moral knowledge is sufficient to give rise to desire). But that would not remove the real problem, namely, since its moral psychology is so sparse, intuitionism gives us no reason to care about morality as much as we do. Knowledge is knowledge, and unless it is shown there is something special about morality, we have no more reason to be concerned with it than with knowledge of any other thing. Without further elaboration, our desire to do what is right appears arbitrary.

Here utilitarianism's account of morality's subject matter is thought to be particularly compelling: we are moved to act on moral requirements because they promote individuals' *well-being*. We are predisposed to commiserate with the pleasures, happiness, pains, and frustrations experienced by others, by way of the natural mechanism Hume called "sympathy" or by inculcated fellow feeling. In either case, it is a fact about us that we have concern for the good of others, and this is normally sufficient to account for our desire to do what is right according to common-sense moral requirements. Moreover, sympathy, once suitably educated and directed to aggregate well-being, can also provide sufficient motive to act on the principle of utility.[17]

The contractualist account of moral motivation differs from both of these views. Unlike intuitionism, the source of moral motivation is not a desire to do what is right just because it is right. It is a desire to act in ways that can be *justified* to others who are similarly motivated on grounds they could not reasonably reject (CU, 116).[18] This "desire for reasonable agreement," Scanlon contends, leads us now to be as concerned as we are about doing the right, fair, or just thing. Scanlon is pointing here to the fact that we do not want others to think, or even have grounds to think, that our conduct is unreasonable or our expectations are extravagant.[19] As Rawls has said, we want to live with others according to terms all would recognize as reasonable, from a perspective they would accept as reasonable. This means we are concerned that our acts and ends be justifiable from a public, and not just a personal, standpoint. This ideal of human relations is appealing to us, and we are drawn to affirm it. This is not to say we always act on this desire, or that we are not often mistaken about the right or most reasonable thing to do. Other sources of motivation compete with it (self-interest and so on), and we are subject to mistaken judgment and self-deception.

---

[16] See Sidgwick, pp 23–5, 34–7, 39f.; W. D. Ross, *The Right and the Good* (New York: Oxford, 1930), pp. 157–60.

[17] See J. S. Mill, *Utilitarianism*, ch. 3.

I take it that what is distinctive about the desire for reasonable agreement Scanlon assigns to us is this: unlike the intuitionist's desire to do what is right just because it is right, this desire presumes an account of the subject matter of morality as part of its content. In this way, it superficially resembles sympathy (construed as a desire for others' well-being). But unlike sympathy, the desire to justify one's actions also presupposes an account of justification (viz., on terms others cannot reasonably reject as a basis for unforced agreement). And that account of justification presupposes a notion of *principles*. The desire for reasonable agreement that accounts for moral motivation is a desire to act only in ways that can be justified to others according to principles they cannot reasonably reject, as a basis for general agreement. As such, it is what we might call a *principle-dependent desire*.[20] Sympathy is not like this; it is a desire for the well-being of particular individuals, or, once corrected, an abstract desire for the sum of well-being. Unlike the contractualist account of motivation, we can describe and understand the object of sympathy without appeal to any moral notions.

---

[18] In "The Significance of Choice" (p. 166), Scanlon defines "the basic moral motivation" as "a desire to regulate one's behavior according to standards that others could not reasonably reject insofar as they, too, were looking for a common set of practical principles. Morality, on this view, is what might be called a system of co-deliberation."

[19] Or as Scanlon says, "People are willing to go to considerable lengths, involving quite heavy sacrifices, in order to avoid admitting the unjustifiability of their actions and institutions" (CU, 117).

[20] Rawls describes the account of motivation in Kant and in "Justice as Fairness" this way. Cf. TJ, sect. 75. Kant defines "respect for the moral law" as "an interest in principles"—*Groundwork of the Metaphysics of Morals*, H. J. Paton, trans. & ed., (New York: Harper, 1964), Ak.401n/69n, Ak.413n/81n.

The significance of this fact will soon become apparent. First, I shall focus on certain features of the desire for reasonable agreement. It is clear Scanlon conceives of it as a primary desire. Like a desire for the good of others, the desire to justify oneself is something people normally want for its own sake. Contrast a Hobbesian contract view, like Gauthier's,[21] which says that moral motives, though real, are conditioned upon the expectation that one will benefit more than if he had no such motives and did not observe moral requirements. Second, as principle-dependent, the desire to act in ways that can be justified is a higher-order desire. It is not a desire for any particular object of action (like the well-being of others or oneself), but is a desire to *regulate* the objects of one's desires and their pursuit in ways (according to principles) none could reasonably reject. These two features imply that having an effective desire of this kind means that one will not (or at least will not want) (a) to impose certain claims and expectations upon others, even if one has a compelling desire to do so, and (b) to do certain things in pursuing one's ends, *because* making those claims and doing those things could not be justified to reasonable individuals.

What does this say about "the nature of the reasons morality provides . . . for those who are concerned with it"? First, it means we have reasons to act in noninstrumental ways, i.e., not for the sake of maximizing anything (satisfactions, well-being, etc.) or effectively promoting a particular end, but for the sake of regulative (moral) principles.[22] This does not mean that acting from reasons of this kind is aimless. Every action must have some kind of object for its end. The ends of action are normally given by the objects of first-order desires (e.g., pleasure, wealth, reputation, the good of another or one's countrymen). Moral desire specifies, not the specific end of action, but the *manner* in which one is to pursue specific ends. One with a higher-order desire to act in ways that can be justified does not want to take any means available (no matter how expedient, pleasurable, or likely) to meet his ends. When this desire is effective, we constrain the means we take to realize our ends, and regulate the aims that are the objects of our first-order desires. Different kinds of motives combine to determine our actions.

Second, the principle dependence of moral motivation implies that the *stringency* of the reasons morality provides is not determined by the strength or intensity of moral motives as compared with that of other desires. Instead, it is determined by the *content* of moral desire. To put this point in its context, it is a feature of Humean accounts that the relative weight of reasons is ultimately determined by the strength of the desires they promote. Where one's reasons conflict (as they inevitably will because of plurality of ends) the balance of reasons is decided, not by reference to these ends, the content of our desires, but by the relative intensity of the primary desires that provide a basis for practical reasoning. Underlying this conception is the basic premise that all desires are, for purposes of practical reasoning, *homogenous:* they are all to be put on a par, no matter what their origin or object, and assessed solely according to their relative strength (as experienced at the time of action, and after reflecting on all one's desires and ascertaining the likely consequences of acting on each).[23] This means that moral motives are to be assigned a weight just like any other desire, and are to be balanced off against them. So, if our moral sentiments are weak or waver under the circumstances, they should count for little in the process of deliberation, and provide reasons that are easily overridden by other concerns. On the Humean account, it would be irrational to assign moral motives, and the reasons they provide, any greater weight than their present strength allows.

---

[21] See *Morals by Agreement*, p. 169.

[22] It may be objected that acting on principles, or the rules they support, is instrumental to our being able to justify ourselves on terms others cannot reasonably reject. But the point is these same principles define the basis for justification and what count as reasons in moral argument. See below.

As an account of duty and obligation, or the weight to be assigned to moral considerations, the Humean account does not accurately represent our practical reasoning. Clearly we do not think that the stringency of moral duties (others' or our own) fluctuate with the current intensity of desire. Reference to desires seems out of place here. In our practical deliberations, we take for granted a background of moral norms. This background manifests itself when doubt arises regarding the legitimacy of our actions or ends: we do not normally ask whether abiding by these norms effectively promotes our prior purposes. Instead, we are concerned to establish whether our ends and proposed actions can be publicly justified to others according to generally acceptable moral criteria. By capturing this aspect of practical reasoning, contractualism (1) indicates what might be wrong with the Humean conception of the weight of moral considerations, and (2) provides an alternative conception of the process of reasoning about our own and others' duties.

(1) The force of the reasons morality provides is determined, not

---

[23] Homogeneity is implicit in views that identify rationality with utility maximization, in the completeness condition: that any two preferences are comparable. See *Morals by Agreement*, pp. 39ff. Scanlon's account of motivation, like Rawls's account of the sense of justice (TJ, 574–5), does not deny that such comparisons can be made, but holds they misrepresent "the nature of the reasons morality provides" (CU, 105).

by intensity of desire, but by the content of the desire for reasonable agreement, i.e., the principles that provide noninstrumental reasons. We have then at least one desire whose object demands that we not act on the overall balance of reasons (as decided by combined intensity of desire). The higher order desire for reasonable agreement is in effect a desire to *moderate* the balance of reasons and to limit first-order desires, as required by principles needed to justify our acts to reasonable persons. Given that content, it follows that acting solely on the balance of reasons settled by intensity of desire could never *satisfy* moral desire (because one's acts could not be justified, especially to those disadvantaged by them, by principles others could not reasonably reject). This means that moral motives, by virtue of their content, have the feature that the reasons they provide *override* the balance of reasons, as determined by the strength of all other desires.[24]

(2) How can this aspect of moral motivation be represented in an account of practical reasoning? From the Humean perspective, for certain reasons always to override all others implies that there are fixed desires whose intensity outweighs the combined intensity of all other desires. To assign to moral motives such an intensity would be highly unrealistic. There must be some other way to represent how moral desire is taken into account in practical reasoning, one that does not involve balancing the reasons it gives against others at all. This means that practical reasoning must have, at least for those with moral sentiments, a different *structure* than that represented by purely instrumental views.

Contractualism implies there are at least two different ways to deliberate about what we have reason to do.[25] These two different methods of reasoning are not exclusive, but supplemental, and one has priority over the other. The first form involves reflecting on one's ends as given by desires and (nonmoral) interests, ranking their priority and making them consistent, and then deciding the most appropriate means (effective, likely, inclusive, pleasurable, etc.) for realizing them. This form of deliberation is *agent-centered,* since it focuses on the present desires and interests of the self and asks what ultimately are my purposes, and what must I do to realize them. However we characterize this account of reasons (take any conception of instrumental rationality), moral sentiments (or interests, as the case may be) are not to be included in determining the balance of reasons. They must be taken into account separately, by a different form of deliberation.

---

[24] This has been challenged (by Foot, Williams, Susan Wolf, and others) on grounds that often prudential concerns or nonmoral ends seem justifiably to excuse us from moral duties. (E.g., it seems appropriate to break certain commitments to take advantage of an exceptional opportunity to further one's career.) But as we shall see, overridingness is a claim about the conclusiveness of the combined set of moral reasons in practical reasoning. There may be occasions where one may breach a specific moral rule to pursue matters of interest, but there must be other moral considerations that warrant such action under those circumstances. Contractualism clarifies when and why these circumstances arise. one may breach a commonly accepted moral duty when one's action under those circumstances accords with a rule that could be justified to others in a hypothetical process of codeliberation.

[25] I am indebted to Rawls for the idea of two stages of reasoning, which correspond to his distinction between reasonable and rational. Cf. Rawls, "Themes in Kant's Moral Philosophy," *Kant's Transcendental Deductions*, Eckhart Forster, ed. (Stanford: University Press, 1989), pp. 81–113, esp. pp 82–8.

This second level of reasoning involves asking whether the kind and balance of reasons given by my agent-centered deliberations satisfy moral requirements, the object of moral desire. According to contractualism, this means asking whether my proposed actions, ends, and agent-centered reasons could be justified on grounds others could reasonably accept. This ultimately involves deciding what principles could be accepted by everyone in a process of ideal codeliberation. Considerations like this often guide us in our individual decisions about what we ought to do. Moral reasoning exercises, in effect, appellate review over our agent-centered deliberations. So construed, moral reasoning *subordinates* the balance of agent-centered reasons.[26]

Arguably, this kind of reasoning is also instrumental, since it is prompted by my moral motives, and involves taking the necessary means (deliberating and acting in a certain way) to satisfy the desire for reasonable agreement. But this is trivial. The important point is that moral deliberation, unlike agent-centered, does not involve reasoning according to principles or precepts that direct that I take the most expedient (i.e., effective, likely, pleasurable, efficient, etc.) means to given ends (e.g., the well-being of others or myself). Instead, it requires appraisal of these ends themselves (to decide if my expectations and claims are legitimate), and assessment of the instrumental means I have *already* decided are appropriate, according to a very different kind of principle: those which others could reasonably accept as a basis for public justification and general agreement.

---

[26] I do not contend that ordinary moral reasoning only involves reasoning about what others could reasonably accept. As Sidgwick says (p. 6), there are several "methods" implicit in ordinary practical reasoning. But this sort of reasoning occupies a significant position in our deliberations. I also do not contend that ordinary reasoning is structured exactly in the way represented in the text. Like the Humean account, the account of different stages of reasoning is an idealization.

What of the further objection, that there may be other higher-order motives to act on regulative principles which also purport to be overriding in the same way. Take Foot's example of principles of etiquette. Many people normally subject their pursuit of ends to these rules as well. How does the higher-order desire to be well-mannered or elegant differ from moral motives?

The content of moral motivation directs that *all* my acts and the content of all other desires be justifiable to everyone on terms they cannot reasonably reject. So it includes any other higher-order desire. This means that the reasons morality provides subordinate all other practical demands in deliberation. Rules of etiquette are not like this. One may breach these rules for moral and other reasons (e.g., to save another, or oneself, from destruction). These excusing conditions are part of the rules of etiquette. Morality also contains excusing conditions (accident, mistake, duress, lack of capacity, etc.). But moral excuses are justifiable only on morality's own terms, i.e., on condition they could be accepted by others as a basis for reasonable agreement. Arguably, rules of law are like this: one may not breach laws for nonlegal reasons, even moral reasons. But to sustain this convention, every legal system professes to be just and to fulfill all moral conditions, hence to be justifiable on terms all could reasonably accept.

So what distinguishes moral motivation from all other higher-order desires to act on regulative principles is that the set of reasons morality provides are *final* and *conclusive* in the context of practical reasoning. There are no higher standards, and no higher procedures, to which those with moral motives might appeal, to assess the system of moral considerations in the course of practical reasoning. Moral deliberation subordinates all other procedures. This is not a matter of choice, but is implicit in the nature of moral reasons. They override, not just instrumental considerations, but the demands of law, custom, and of social rules in general. In this sense, the desire for reasonable agreement is a *highest-order desire*.[27] It is in effect a desire to subject the reasons that result from all other processes of deliberation to the test of a final and conclusive procedure.

If we see the reasons morality provides as involving a different form of practical reasoning, one that assesses and subordinates the balance of instrumental reasons and the demands of all other forms of deliberation, then we have an idea how moral considerations offer reasons of a special kind. Those who are morally motivated have overriding reasons of a noninstrumental kind to act on moral requirements, *whatever else* their desires and interests may be. This does not mean moral principles are strictly "categorical." But contractualism indicates something of significance in Kant's notion.

---

[27] Cf. Rawls, "Kantian Constructivism," p. 533, on the sense of justice as a "highest-order desire." Suppose someone (e.g., Kierkegaard's aesthete) resolves never to breach rules of elegance; they override all other motives, even moral ones (e.g., he will not save a drowning child for fear of wetting his white suit). This person regards rules of etiquette as if they were moral rules. There is nothing about etiquette that would account for the special stringency he assigns to it. We can only account for it by the strength of this peculiar desire. Also, it is doubtful this person can even be ascribed the basic moral desire, that *all* his acts be reasonably justifiable.

III. THE REASON DEPENDENCE OF MORAL MOTIVATION

I consider now how contractualism can elucidate another feature of Kant's view, the idea that moral motivation is based in reason. Kant speaks of an "intellectual cause" of moral motives, and of reason "directly determining the will."[28] This is often construed as a causal claim about the origins of moral motivation: by rational deliberation alone, reasoning in accordance with certain principles, one without moral motives (the amoralist) can come to have an effective desire to act on moral requirements for their own sake. Whether this is a correct reading of Kant is not my concern. There is yet another way to construe the reason dependence of moral motivation.

I have argued, in effect, that having moral sentiments implies one's motivations have a certain *structure,* a hierarchically ordered complex of desires. Where does this structure come from? According to contractualism, practical deliberation involves: first, ascertaining the balance of agent-centered reasons and the actions needed to effectuate them; then, reflection on our proposed actions, desires, and agent-centered reasons, to see if they accord with principles that could be reasonably accepted by everyone in a process of ideal codeliberation. Moral motivation in turn is defined as a desire to act only in ways and according to reasons that could be justified by this hypothetical procedure. Having this highest-order desire implies that, whatever my agent-centered reasons, I have a willingness to subject them to a higher procedure, and a desire to act on its outcomes by regulating means and even modifying the ends that direct my agent-centered deliberations.

This allows the inference that the complex of motivations has a formal basis in practical reasoning. A hierarchy of motives (first-order . . . . highest-order desires) is imposed by the stages of reasoning we go through to decide what we ought to do. The structure of practical reason is, in effect, reflected in our desires. This structure thereby influences the content of our agent-centered desires—we often revise our ends if they do not accord with moral principles

---

[28] *Critique of Practical Reason,* Ak 79, also *Groundwork,* Ak.122n/Paton, 128n.

—and so contributes to the generation of agent-centered reasons and actions done for those reasons.

That moral deliberation imposes a formal structure on our motivations, thereby regulating our desires and actions, is a central feature of Kantian views.[29] (It is part of what Rawls means by "the priority of the right over the good.") The complex of motivations has this structure due to the mediating influence of moral desire. Moral desire itself is reason-dependent in a more direct way. According to contractualism, moral motivation is ultimately *principle-dependent;* it rests on a desire to act on principles that could be justified to reasonable individuals. To say this desire is principle-dependent means, first, we have this desire for a reason; moral motivation is not an arbitrary taste. It also means the reasons we have for moral motivations cannot be accounted for by prior and independent desires, like Humean sympathy or love of mankind. By contrast, contractualism implies that the reasons that support our moral motives are the same reasons that underlie our moral judgments. We can explain why we have these motives, and why we want to act on moral requirements, only by appealing to the content of our deliberations about moral issues, viz., our judgments about what acts could be justified by principles others could reasonably accept.

Now granted, we would not have moral motives were it not for moral education. We acquire moral motives over a long time, acting in our youth initially out of authority, often times fear of punishment, and later from a desire to please those we love or respect.[30] To say moral motives are reason-dependent does not imply a special causal process, through which the amoralist could acquire moral desires purely by intuition of or reflection on sound moral principles. But we must distinguish the conditions of the *origins* of moral motives in other desires, from the conditions needed to *sustain* our moral motives *once acquired.* Once one has reached maturity and his reasoning abilities are fully developed, what sustains moral motivation is not some other desire (like fear of punishment, or sympathy). It is the judgment and conviction that one's acts accord with principles: those which could be justified to reasonable individuals in a process of codeliberation. This desire is *reason-dependent*, because it depends on exercising our capacities for practical reasoning in a certain way. It requires an understanding of moral rules and an ability to apply them in judgment. And ultimately it involves an ability to reason about what actions and institutions could be justified to others by principles they could reasonably accept. So moral motivation is not like sympathy. It is not a desire or predisposition we naturally have which could be sustained *independently* of judgment and the exercise of our capacities for practical reasoning. Whatever role sympathy might have in accounting for the origins of moral motives, it is not what sustains them once they exist.

---

[29] More than a hierarchy of motives is required for a Kantian view. Harry Frankfurt [in "Freedom of the Will and the Concept of a Person," this JOURNAL, LXVIII, 1 (January 14, 1971): 5–20] says we have higher-order volitions to have certain desires and to be a specific kind of person, which regulate our first-order desires. But there is nothing specifically moral about these desires as Frankfurt describes them. They might as well be agent-centered (e.g., the desire to be elegant). Moreover, from where do these higher-order desires come? For Frankfurt, they ultimately stem from a kind of radical choice. Kantian accounts differ on both issues. More than being a desire to have certain desires, moral motivation is principle-dependent. And these principles, we shall see, are not based in radical choice, but in our capacities for practical reasoning.

[30] Cf. TJ, ch. 8, on development of the sense of justice; CU, 117

A third way to describe the reason dependence of moral motives stems from our beliefs regarding the *status* of our moral convictions. For most of us, it is our belief that our moral judgments are sound which moves us to act on them. I would no longer want to abide by moral demands under all the conditions I do, did I not believe that my judgments and actions are supported by true or reasonable principles. But my belief that there are such principles is, according to contractualism, just my belief that there are principles that could be justified to others in an ideal process of codeliberation. If so, then my desire to act on these principles is sustained by the conviction that my moral deliberations are not in vain, but are answerable by this objective criterion, knowable to me through practical reasoning.

Arguably, it is misleading to say that belief that a judgment is true is reason-dependent. Instead, our beliefs are dependent (causally, let us say) on certain facts that obtain in the world. But consider again the contractualist account of moral truth. The "grounds of truth" are not any specifiable order of facts, natural or non-natural, which exist prior to and independently of human association and reasoning, and which are imposed upon us via observation or moral experience. Instead, it is the principles that *could* be agreed to in an ideal process of codeliberation. This is not a fact independent of human reasoning. It is a construction articulated by principles—an Idea of Reason in Kant's sense—based in the common reasoning of mankind, as we go about exercising our practical capacities to decide how to live together socially. We ultimately appeal to this construction to guide as well as to justify our judgments about what we ought to do. It also sustains moral motivation.

I have mentioned three ways moral motivation has a basis in reason, as opposed to having its basis in independently comprehensible natural desire (like sympathy): (1) the structure of our motivations reflects the structure of practical reason; (2) the content of moral motives themselves, of these highest-order desires, is supplied by ethical principles that govern our moral deliberations; (3) our acceptance that these principles are true, i.e., are justifiable as reasonable to others similarly motivated, maintains our moral motives. Still, is this sufficient for Kant's purposes, to define moral autonomy? Nothing has been said yet about the bases for ideal codeliberation and reasonable agreement on which we as individuals deliberate and want to act. Here we must look to the structure and details of Scanlon's hypothetical agreement. Scanlon says little about this in his initial article. He says it is unlikely that the principle of average utility could be justified as reasonable on a contractualist view. But suppose it could; then the sense in which moral motives have a basis in reason would be attentuated, since the ultimate principle practical reasoning provides would direct that we maximize average well-being. To avoid this problem (a heteronomous motivation on Kant's view) we need further specification of the ideal procedure of codeliberation.

In more recent work, Scanlon says that the moral principles contractualist agreement seeks are "for the [self] regulation of behavior," an idea which involves two components: a specifically moral capacity, "the ability to reason about what could be justified to others"; and a nonmoral "capacity for critically reflective, rational self-governance."[31] These are the basic capacities for practical reasoning. They enable individuals to formulate ends and to pursue them effectively, and to regulate their conduct within social groups. As such, these capacities provide the basis for persons' being the primary units of agency and responsibility, and their conceptions of themselves as free. Moreover, these capacities are conditions of our practical deliberations, both reasoning about our purposes, and understanding and acting on the results of the hypothetical codeliberations contractualism defines (cf. SC, 175). They are the practical powers that enable individuals to have moral motives, and to take an interest in, and act on and from, the codeliberation to moral principles which defines the subject matter of morality.

Suppose we construe this ideal codeliberation as a procedure that incorporates essential features of these capacities. This interprets contractualism as a version of Rawls's "Kantian Constructivism." According to constructivism: (1) first principles of right and justice, as statements about good reasons, are regarded as true or false by reference to the social role of these principles in providing a public basis for justification in an (ideal) society of reasonable and rational persons (a well-ordered society in Rawls, and the Realm of Ends in

---

[31] "The Significance of Choice," pp 173–6.

Kant). Moreover, (2) these principles are specified by a procedure of construction, the form and structure of which express our free moral personality, and therewith essential features of our capacities for practical reasoning.[32] Scanlon's contractualism satisfies (1). But (1) alone does not, I believe, clearly ensure that the principles to which Scanlon's parties agree will not be based in considerations of individual well-being. If not, contractualism is just a more roundabout justification for utilitarianism. This presents a challenge to Scanlon's view. One way to avoid this problem is to structure the process of codeliberation so that a description of the two practical capacities enters into the reasoning by which moral principles are justified. Principles resulting from that procedure could then be construed as an *expression* of our practical conception of ourselves as free, and as reasonable and rational beings (cf. TJ, 252, 505).

If we see moral motivation in this context, we get a clearer sense of how moral motives might have their ultimate basis in practical reason. The principles that result from a constructivist procedure of codeliberation would be what we ultimately are trying to reflect on in our individual moral deliberations. So our desire to act only in ways that could be justified as reasonable can be construed as a desire to act on principles that are the result of a procedure designed to express our practical nature as free moral beings.[33] By acting on these principles and from these motives, we act for the sake of principles "given" to us by practical reason out of itself.

Kantian constructivism is a complicated subject I shall not discuss further. Whether Scanlon sees contractualism this way is a separate issue. My point is that a constructivist interpretation fits the primary aspects of contractualism, and provides a way to carry through one of Scanlon's primary aims: to show that the reasons morality provides, if not strictly categorical in Kant's sense, are at least of a special kind.

IV. CONCLUSION

To sum up sections II and III, contractualism indicates how morality can provide (a) conclusive reasons for acting, (b) of a noninstrumental kind, (c) which have a basis in practical reasoning. And it does this in a way that is compatible with the assumption that (moral) desire is a condition of having (agent-centered) reasons to act on moral requirements. But while it satisfies this condition, contractualism denies three basic claims of Humean accounts: (1) the weight of moral reasons is given by the intensity of moral (or other) desire; (2) the content of moral reasons (normative principles) is specified instrumentally, by what most effectively promotes prior and independent ends; and (3) reason imposes no substantive requirements on desire or action. To assume desire is a condition for having reasons does not require accepting these conditions.

---

[32] See Rawls, "Themes in Kant's Moral Philosophy," pp. 95–7.

[33] Cf. TJ, 572: "The desire to act justly and the desire to express our nature as free moral persons turn out to specify what is practically speaking the same desire." See also TJ, 256.

I have not argued that moral demands are "inescapable," or that moral reasons apply to everyone no matter what his desires and interests. Scanlon mentions the unavoidability of moral reasons, but does not address it (CU, 105). It raises the question whether there are what Thomas Nagel calls "objective (moral) reasons." One condition for the objectivity of moral reasons is that the soundness of moral principles be independent of the desires of actual individuals. Contractualism says the "grounds of truth" for moral judgment are the principles that could be agreed to by members of an ideal community, each of whom has a desire to find principles for regulating actions and institutions which will serve as a basis for public agreement. Whether there are principles meeting this description does not depend on the desires *we* now have, or even on our beliefs and conventions, and the agreements we might come to on these bases. It depends on the knowledge, deliberations, and motivations of these hypothetical persons. By abstracting from these features of our actual situation, the interpersonal perspective defined by contractualism provides an objective point of view. Our moral judgments are then true when they accord with principles that could be unanimously agreed to from this objective perspective.

This provides a necessary basis for objective reasons. Is it sufficient? Arguably, since morality affects not just belief but also the will, it must be shown how we can be motivated to act on moral demands. Here Humeans balk at the idea that morality provides reasons that apply to everyone. For how can awareness of moral requirements, even assuming they are true, move me to act in the absence of antecedent desire?

Even if we accept the Humean thesis that desire is a condition of the applicability of reasons, this condition can apply at most to agent-centered reasons. Moral reasons, by contractualist definition, cannot depend for their validity upon *our* antecedent desires. Instead, they explain *why* we have certain desires. Or as Nagel[34] says, they explain how moral motivation is possible. I have argued in effect that a constructivist contractualist account provides a way to approach that issue. What bearing it has on the broader unavoidability issue is not something I can discuss further here.

SAMUEL FREEMAN

University of Pennsylvania

---

[34] *The Possibility of Altruism* (New York: Oxford, 1970), pp. 12–5

COMMENTS AND CRITICISM

THE CAUSAL HISTORY OF COMPUTATIONAL ACTIVITY: MAUDLIN AND OLYMPIA*

Tim Maudlin has, in a recent paper,[1] constructed a thought experiment that raises puzzling and important questions about what it is we attribute to a physical system when we describe it as performing a computation. His conclusion is that this thought experiment repudiates the possibility of a computationalist theory of consciousness. I shall argue that it would be more plausible to adopt instead the following conclusions: (1) the notion of "computing" is, like many other cognitive notions, best characterized in terms of a causal theory that asserts that computational activity is essentially the product of a certain causal history, and (2) the causal history essential to computation has a particular structure, one in which the computed input plays a peculiar role vis-à-vis the activity it induces in the computing system. The spirit of conclusion (2), moreover, is well applied to a wide variety of other cognitive activities, it seems to me, and thus illuminates something fundamental about the sort of causal interaction between a cognizing subject and a cognized object which is generally essential for many types of cognition.

Let us first introduce a conceptual tool that will be applied later. The tool consists of a distinction between two sorts of causes of physical events. Consider, e.g., the arrival of a letter in one's mail box. The causes of this event could be described in terms of, say, the various forces applied to the letter which induced the letter's motion from its point of origin to the mail box. I shall call such causes *active causes,* which work in general to cause events to occur to an object *X* by the net deployment of forces, or the transmission of energy, or the application of stimulants to or against *X*, perhaps by way of intermediate causal chains (supplied, e.g., by a medium through which active causes are transmitted). I shall distinguish active causes from *passive causes,* which work in general by the net nondeployment of forces (or energy or stimulants). A passive cause is nothing more than the nonoccurrence of an intervening event; a passive cause of the letter's arrival is that there was no insurmountable

---

* In the course of writing this note, I have profited from a number of discussions with John Winnie, though any errors in my analysis are certainly mine alone.

[1] "Computation and Consciousness," this JOURNAL, LXXXVI, 8 (August 1989): 407–32.

0022-362X/91/8806/304–16

barrier that deployed a force against the letter blocking the letter's motion. To say of some state of affairs that it works as a passive cause of an event that occurs to an object *X* is just to say that this state of affairs is such that it does not block or obstruct the active causes that induce the *X*-related event. The fact that the Granville police did not arrest me and throw me in jail this morning worked as a passive cause of my afternoon stroll to the Student Union, since this nonarrest allowed my stroll to transpire. This active/passive cause distinction is admittedly somewhat vague, but it is sufficiently precise for the purposes at hand.

I MAUDLIN ON COMPUTATION AND GAA THEORIES OF COGNITION

What must a physical system do, and what must its structure be, if it is to be aptly described as a Turing machine running some algorithm π on some input τ? Such a system must contain, at least, an arbitrarily long tape and machinery that can manipulate the tape's contents. This machinery must, moreover, produce the correct output for a running of π on τ by following the "algorithmic rules" as laid down by π. A Turing machine's rule-following behavior consists of three elements: (1) the machine must write, at each step in the computation, either a 0 or a 1 at its current tape address, depending on the directives of π for the machine's current state and the original content of the current tape address. (2) The machine's read/write head must move, at each step, one address to the right or left on the tape, depending on π's stipulations, and (3) the machine must enter, at each step, the appropriate final state, again in accordance with π. To say of a physical system that it is currently running π on τ, then, is to say, first, that the system has been given τ as its input and that it is executing the appropriate sequence of write/move/change state operations (hereafter, W/M/CS operations) for an application of π to this input. Second, it must also be true that, had the system been given an input θ not equal to τ, the machine would have executed the appropriate W/M/CS sequence for θ. These two conditions combine to assert that the activity of the computational system is generally appropriate for all (actual and possible) inputs. I deem this latter condition the ***generally appropriate activity condition for computation*** (hereafter, the GAA condition for computation). Oversimplified, Maudlin's view takes the GAA condition for computation to describe what we attribute to a physical system when we describe it as computing π on τ.[2]

---

[2] Maudlin adds additional restrictions on computational systems to handle the problem posed by the "rock instantiation" (cf. p. 418), but this formulation of his theory of computation suffices for all present purposes.

It is interesting to note that GAA type conditions seem to work well to characterize a wide variety of cognitive activities. Consider, e.g., a GAA characterization for reading: to say of a subject *S* that it is reading a text *T* is to say, first, that *S*'s current verbal representation is (more or less) identical to the content of the text *T* he is currently presented with, and second, that had *S* been presented with some text *T*′ not the same as *T*, *S*'s verbal representation would have (more or less) coincided with the content of *T*′. That is, to say of *S* that he is reading *T* is to assert that the architecture of the subject/text system is such that the verbal activity of the subject will generally be appropriate given the content of the given text, for the actual text *T* or any possible (readable) text. Or consider a GAA condition for seeing: to say of *S* that he is seeing object *O* is to say, first, that *S*'s current visual experience more or less coincides with the actual appearance of *O*, and, second, that, had some object *O*′ not the same as *O* been present to *S*, *S* would have had a visual image more or less identical to the actual appearance of *O*′. The list of GAA conditions for cognitive activities could easily be extended to cover other cognitive activities, such as "referring," "knowing," "perceiving," etc.[3]

It is not surprising that GAA conditions seem to offer plausible analyses of cognitive activities. Such activities generally involve a subject (person or machine), who does the cognizing, and an object, which is somehow "cognized" in virtue of the subject's activity. (In the case of computation, the cognizing subject is just the computational machinery (minus the tape), and the cognized object is the input given initially in the tape contents.) It is frequently proclaimed that it is just in virtue of the fact that the subject's activity is causally attributable in some appropriate way to some object *O* that the subject's activity amounts to a cognitive act performed on *O* (and no other object). Now the generic notion of "causation" is often explicated in terms of the truth value of counterfactual statements, hence it has become rather standard to explicate the relevant causal connection between a cognizing subject and the cognized object in terms of counterfactual conditions asserting that, not only is the subject's activity appropriate for the actual object presented to the subject, but the subject's activity would also have been appropriate for some other object, had this other object been presented instead. This sort of condition is precisely what GAA conditions stipulate for cognitive activities in stipulating that the subject's activity be generally appropriate for any suitable object.

---

[3] I have no idea whether Maudlin himself would countenance the generalization of GAA characterizations to cognitive activities other than computation; I merely claim that such a generalization seems rather plausible, given certain simple intuitions about the nature of cognitive activity.

II. THE RIGID AND SILENT OLYMPIA

Maudlin's thought experiment reflects on the architecture and activity of Olympia, a candidate Turing machine. Olympia's tape consists of a sequence of water troughs, which individually record the data entry 0 or 1 by being either empty or full of water. One of Olympia's components is "Klara," an ordinary Turing machine programmed to run π. Another component is an armature, which is little more than a rod on wheels with some additional apparatus. If Klara were given τ as an input on the tape, she would perform a possibly long and complex sequence of W/M/CS operations before, in the end, producing the correct output for a running of π on τ. Now the armature is rigged with a water hose, and barriers, levers, and drain plugs are attached to some of the troughs, so that, as the armature runs on its wheels along the troughs, it mechanically empties and fills the troughs just as Klara would have had Klara run π on τ. Furthermore, by a clever trick of defining machine states "disjunctively," Maudlin ensures that the passage of the armature along the troughs amounts to Olympia's passage through the same sequence of machine states that Klara would have passed through had Klara performed this computation. In sum, the armature is rigged so as to mimic the W/M/CS operation sequence that Klara would perform in running π on τ. But the armature, unlike Klara, is merely executing a rigidly predetermined sequence of such operations. The only thing the armature can do is perform this unique sequence of operations. The activity of the armature is, in this sense, dumb. The armature's activity, clearly, cannot by itself do the work of running π on τ, despite the fact that it executes the appropriate sequence of W/M/CS operations for this computation. For given the armature's rigid structure, it is not true that the armature would have performed the appropriate sequence of W/M/CS operations had it been given any input θ not equal to τ. That is, the GAA condition for computation is not met by the armature's activity.

Maudlin's important insight is that the GAA condition can be met by supplementing the armature with machinery which essentially does nothing, i.e., which engages in no "activity" (according to a vague but common interpretation of this term) whatsoever when the input is τ. Maudlin imagines copies of Klara attached to floats; one such float is fixed in place in each trough. As the armature passes over each trough, the trough's float is released so as to "detect" the content of the trough. If the content of the trough should differ from what it will be if the initial input is τ, the trough's float will rise or fall, thereby stopping the armature's activity and causing the attached copy of Klara to be pulled into action, whereupon this copy of Klara will complete the W/M/CS operations appropriate for the given input just as an ordinary Turing machine would. If the content of the tape address is what it should be when the input is τ, the floats do not move at all, and hence the chains connecting the floats to the copies of Klara just lay slack, and the armature performs its work unhindered. The presence of these copies of Klara, along with the floats and chains, serves to support the required counterfactual condition: we can now say with confidence that, if any input θ not equal to τ were given on Olympia's tape, Olympia (construed now as the armature and copies of Klara working together) would execute the uniquely appropriate W/M/CS sequence for θ. But again, as long as the input is τ, no float moves, hence no copy of Klara is ever called into play, and thus this extra machinery does nothing. Olympia's entire activity on τ consists of the armature's operations as it runs along the troughs. Despite the intuition that this activity is "dumb," there now appears to be a serious case for the claim that Olympia succeeds in running π on τ just in virtue of this activity, along with the (inert) copies of Klara in the background, doing nothing but supporting the relevant counterfactuals. Undeniably, Olympia meets the GAA condition for computation, and surely, Maudlin plausibly claims, this is all that is required for a system to run π on τ.

If we grant that Olympia succeeds in performing this computation when given input τ, then it seems that we must also grant that Maudlin has succeeded in refuting a computationalist theory of consciousness. Such a theory, Maudlin conjectures, would claim that (A) it is sufficient for a system to be in some conscious state ψ that the system run some program π on some input τ. Also, such a theory would hold that (B) for a system to be in any conscious state, it must be describable as a nontrivial Turing machine running an algorithm on some input. Finally, any theory of consciousness will surely endorse the physicalist claim that (C) distinct systems that are engaged in identical sorts of activity will be in identical conscious states (if any). (C) is intended to ground the intuition that the presence or absence of any object that is causally disconnected from a system's activity cannot make a difference with respect to the system's status as conscious (or with respect to the quality of the system's conscious state, if it is conscious). Maudlin's purpose in constructing Olympia is to demonstrate the joint inconsistency of (A), (B), and (C). It follows from (A) that Olympia must be in conscious state ψ when she runs π on τ. But suppose we insert a "block" in between the gears of each of the Klaras' mechanisms such that these mechanisms are disabled and can no longer do the work of supporting the relevant counterfactuals. The activity of Olympia on τ (the armature's activity) is now a noncomputation, since the GAA condition for computation is no longer met. But the blocks can be inserted in a way so as not to interact causally with Olympia's activity in any way. They need not even be touching Klara's gears or engaged in any sort of interaction with any part of Olympia whatsoever. Given (C), then, the presence of the blocks cannot change Olympia's conscious state in any way: it must still be in conscious state ψ with the blocks in place if it was in state ψ without the blocks. But given (B), the blocked Olympia can no longer be conscious at all (and thus, of course, cannot be in state ψ) given the blocked Olympia's noncomputational status. The computationalist theory of consciousness that consists of the conjunction of (A), (B), and (C) is thus reduced to absurdity.

If it is true that Olympia is a legitimate Turing machine, and that she does indeed run π on τ in virtue of the fact that she grounds the GAA condition, it must nonetheless be granted that she is a decidedly unusual one. One of the most unusual facts about Olympia concerns the causal history of her activity when given the input τ. This causal history is such that Olympia will perform the correct W/M/CS sequence for this input even if her tape sensitivity is deleted, say, by removing the floats from the troughs and letting the armature make its run. This contrasts sharply with the causal history of Klara's (or any ordinary Turing machine's) activity when she runs π on τ, since if Klara were to lose her tape sensitivity (say, by deleting the reading component of her read/write head) she would almost certainly not proceed to execute the presumably long and complex W/M/CS sequence appropriate for a running of π on τ. I suggest that this difference between Olympia and Klara, the difference in the causal history of their "computational" behavior, bears close scrutiny.

The simplest method of rendering the Olympia/Klara distinction, it seems to me, is in terms of an active/passive distinction in the causal histories of their respective activities on τ. In short, the list of active causes of Klara's "π on τ" activity includes the informational content of the tape. More broadly, the "reading process" that gives Klara her tape sensitivity functions as an active cause of Klara's "π on τ" activity. This reading process consists, of course, of a sequence of "reading acts" which are individually performed at each step in the computation, and which enable Klara to react appropriately, given the content of her current tape address; however Klara's reading mechanism is structured, and however she executes this process, the following is almost certainly true:[4] the physical event or process that constitutes, in Klara, a reading of a 1 at a tape address cannot be the same type of physical event or process that constitutes a reading of a 0 at a tape address. If this were not true, Klara's reading mechanism could not distinguish between a 0-entry and a 1-entry, and thus Klara would lack the tape sensitivity required to run a computation. Putting the matter loosely, a 0-reading must consist of the transmission of a signal, presumably in the form of the application of a particular sort of force (or energy, or stimulant) from a tape address through the read head and into Klara's machinery, and a 1-reading must consist of the transmission of a distinct signal in the form of a sort of force (or energy, or stimulant) not identical to the 0-type force.[5] Klara's entire reading process consists, therefore, of a pattern of sequentially deployed forces exerted by the tape on Klara's computational machinery, a pattern of forces that contains the tape's informational content in coded form. These forces must surely be among the active causes of Klara's activity when she runs π on τ; without these forces, Klara lacks the needed impetus to execute the appropriate W/M/CS sequence.

In contrast, Olympia's reading process, and thus the content of her tape, is a merely passive cause of Olympia's "π on τ" activity. That is, the fact that the input τ is given on Olympia's tape simply results in the fact that the armature is "allowed" to perform its predetermined W/M/CS sequence. Since no floats move on this input, no copy of Klara is activated, and the armature's activity merely proceeds without intervention. Note, e.g., that Olympia's reading process on τ consists entirely of the troughs' chains simply laying slack, thereby conveying the information that the given input is τ. Hence the physical process that constitutes a 0-reading by Olympia is, amazingly enough, physically identical to the physical process that constitutes a 1-reading (both processes are constituted by a trough's chain remaining slack). But whereas in an ordinary Turing machine this fact would mean that the requisite tape sensitivity was simply missing, this fact does not, apparently, cause a problem for Olympia, given that she manages to execute the correct W/M/CS activity despite this fact. The reason for this is just that all the active causes sufficient for Olympia's "π on τ" activity are installed in the interior of Olympia's computational machinery; these active causes consist of the force that pulls the armature down the troughs, together with the forces exerted by the mechanical riggings (the hose, barriers, levers, and drainage plugs) that collectively perform Olympia's "π on τ" work. It is, I believe, the internality of these active forces, and the fact that these active forces do not trace down to the informational content of the tape through the tape reading process, that makes the claim that Olympia is computing the input given on the tape at least a little perplexing. But it is not at all obvious that the tape's content must function as an active cause of the Turing machine's W/M/CS activity in order for that activity to be legitimately construed as a computation of the input given on the tape. We investigate this issue in the subsequent section.

---

[4] The only exception would be if Klara herself had an Olympia-like architecture; here, I assume her architecture to be that of a garden variety Turing machine.

[5] Alternately, a 0-reading could consist of the nontransmission of any force, while a 1-reading could consist of a transmission of a nonzero force. While a 0-reading would thus amount to a kind of "passive" tape/machine interaction, it would nonetheless not be true that the entire reading process were passive, i.e., consisted of no force deployments against the read head. It is only reading processes which are entirely passive which I critique below.

III DREAMER/BELL SYSTEMS AND COGNITION

Let us imagine the following scenario: a subject *S* has a piece of paper *T* before his eyes (or any ocular organ) on which the sentence *p* is written. *S*'s eyes are directed sequentially to each word of this sentence from left to right, and, as *S*'s eyes are pointed toward a particular word, *S* simultaneously pronounces this word to himself. Finally, let us add the assumption that *S* is asleep and that his verbal activity is merely the product of a purely auditory dream. It is simply a coincidence, in other words, that the content of *S*'s auditory dream happens to coincide with the content of *T*. Clearly, despite this coincidence, *S* is not reading *T*. This is because there is virtually no causal connection whatsoever between *T*'s content and *S*'s verbal activity, but the activity of "reading a text" surely requires some such causal connection. This is just to say, of course, that the GAA condition for reading is not met by this subject/text system.

Let us now "Olympia-ize" this example in the following way: in conjunction with our dreamer *S*, whose eyes trace the words in *T* while privately stating each word while his eyes point toward it, we rig up a "bell system." The bell system is designed so that, if the text before *S*'s eyes is just *T* (and thus contains just the sentence *p*), the bell does not ring. But if the text were to have been different, such that at any point the word *S*'s eyes point toward does not coincide with *S*'s inwardly pronounced word, the bell will ring at precisely this point, thereby waking *S* up. *S* will then proceed to read the remaining given text in the usual way, using the portion of his mind normally deployed by *S* when *S* is awake and reading. The presence of the bell system serves to guarantee that the GAA condition for reading is met by the dreamer/bell system: this system's verbal activity will generally be appropriate for the given text, whether the text is *T* or any other text, since, should the dream-induced verbal activity fail at any point to be appropriate for the text, the bell will ring, enabling the system to read in the normal way whatever text happens to be present.

Let us reflect now on *S*, his eyes studying *T*, dreaming his auditory dream, with the bell system in place. The bell will not ring, naturally, since the given text is *T*, and *S*'s verbal dream coincides precisely with *p*. The interesting, albeit odd, question we must now consider is whether the dreamer/bell system, conceived as a single unit, is reading *T*. The case for the claim that this system is reading *T* consists, apparently, entirely of the fact that it meets the GAA condition for reading. But I think there are rather compelling reasons for denying that this system is reading *T*, and thus for denying that the GAA condition for reading is in fact sufficient for reading. The principle problem with the claim that the dreamer/bell system is reading *T* concerns the causal history of the system's verbal activity. In particular, this causal history does not consist of the stimulation of the system's sensory organs by the text. (The system's sensory organs will include *S*'s sensory organs together with the bell system, itself clearly a sensory organ of some sort.) To the contrary, this sensory physiology is completely dormant throughout the course of the system's verbal activity. This is to say that the text does not function as an active cause of the dreamer/bell system's verbal activity. The presence of text *T* simply works as a passive cause of the verbal activity, since this presence does nothing relevant to the verbal activity but prevent the bell from ringing and thereby interrupting the dreaming process. The active causes of this activity will be the sort of causes that induce dreams, such as a brain-wave pattern characteristic of REM sleep, or perhaps the secretion of a chemical in the brain which is produced only in sleep and works to generate hallucination-like experiences. To say of any experience that it is a dream is just to say, it seems to me, that the experience is produced by just such a causal history, viz., a causal history that traces down to active causes entirely internal to the dreamer's own cognitive machinery.[6]

---

[6] This characterization is incomplete, clearly, in that it fails to distinguish dreams from other experiences whose active causes may be internal to the subject's own cognitive machinery, e.g., hallucinations or acts of the imagination.

Of course, the external world plays a role in the causal history of a dream, too, insofar as the world external to the dreamer's mind serves not to interfere with the ongoing dreaming process, say, by not waking the subject up. A causal history of a dream experience will locate the active causes of the experience in the interior of the dreamer's mind, and usually ascribe to the external world a purely passive causal role.[7] But this is the very type of causal history which generates the dreamer/bell system's verbal activity in our scenario, hence it seems quite clear that this activity is as much a dream vis-à-vis the dreamer/bell system as it is vis-à-vis the dreamer alone. The verbal activity of this system is not, I am suggesting, tied sufficiently tightly to the text to establish the activity as a reading of the text. It is, I claim, only in virtue of the fact that a text *T* functions as an active cause of a subject's verbal activity that it is correct to say of the subject's activity that it is a reading of *T*.

The point I am making concerns, in a clear way, the meaning of the term 'reading'. In short, I claim that "reading" is an activity a subject performs to a text with the use of the subject's sensory organs, i.e., by the active stimulation of these organs by a pattern of forces deployed, so to speak, by the text (usually by way of a medium like light rays). This pattern contains the content of the text in coded form, so that the subject can apprehend the text's content by interpreting the stimulated state of his sensory organs. One cannot read a text, I am claiming, when one's sensory organs are dormant and unstimulated any more than one can ride a bicycle when the muscles of one's limbs are dormant and unstimulated. The causally active role of the text in the causal history of the subject's verbal activity is simply essential for the subject to be reading the text; where this condition is not met, the subject is not reading the text even when the GAA condition for reading is met by this subject, and thus even when the subject's verbal activity is identical to what it would have been had the subject actually read the text. This is the moral of the dreamer/bell system for reading.

Dreamer/bell systems, it is interesting to note, are easy to construct for a wide variety of cognitive activities, and all such thought experiments seem to me to elicit the intuition that the system is not engaged in the sort of cognitive activity at issue. Consider a dreamer/bell system for "referring": our dreaming subject *S* sits before a table with a candle, and simultaneously dreams an image of a qualitatively identical candle. The bell system is rigged to ring only if the appearance of the table's actual contents fails to be qualitatively identical to *S*'s dreamt visual image, so of course the bell does not ring. (Were the bell to ring, *S* would wake up and study the table's contents in the usual way; he would then attribute some property to whatever he observed on the table, and this reference to the table's contents would constitute an ordinary referential act.) *S* now performs, in his dream, the intensional act of attributing to the candle (the candle he is "observing") the property of being bright. The relevant question, of course, is whether the dreamer/bell system has hereby actually referred to the real candle on the table, given (1) that the system's phenomonological state (consisting of the visual experience of a candle and the attributive act) is identical to what it would have been had the dreamer been awake and thus referred to the real candle in an ordinary way, and (2) that, had the table's contents been different in any way, the system would have engaged in a phenomenological state that would count as an ordinary (perfectly legitimate) act of reference to whatever happened to be on the table. The most reasonable answer, surely, is that the system has failed to refer to the real candle, given the nature of the causal chain that induces the system's intentional act. The deviant nature of this chain, again, consists of the fact that the active causes of the referential act are intrinsic to the subject's dreaming process, and do not in any way trace down to the real candle. The real candle is merely a passive cause of the referential act, in that its presence simply prevents the bell from ringing and halting the dream-related active causes that generate the referential act. But the real candle's passive causal role is not of the sort that will connect the real candle to the referential act in a way that establishes this act as a reference to this candle. At most, the dreamer/bell system has referred to the dreamt candle, and not the real candle at all. The point here, obviously, is the same as for the reading scenario: it is necessary that an object *O* function as an active cause of a "referential act" in order for that act to be a reference to *O*. Similar thought experiments could be given for "perceiving," "seeing," "believing," "knowing," and many other cognitive activities. In general, for any or almost any cognitive activity, it is necessary that the cognized object play an active role in the causal history of the cognizer's activity for the cognizer to be actually engaged in "cognizing" the object; I hereby label this condition for any cognitive activity *X* as the *causally active*

---

[7] Sometimes the external world plays an active role in the quality of a dream's content, if externally supplied stimuli are "assimilated" into the dream content without interrupting the dreaming process (when, e g., a subject dreams of being attacked by a tiger and wakes up with a cat scratch on his face). My point is made here by the example of dreams in which no such assimilation takes place.

object condition for X (hereafter, the CAO condition for *X*). When the CAO condition is not met by a subject/object system, there is no reason to describe the subject's activity as a cognitive operation performed on the object. In my view, it is surprising but not paradoxical that a subject/object system may meet the GAA condition for a cognitive activity without meeting the CAO condition for that activity. Such systems are instantiated by dreamer/bell systems of the sort considered above. The GAA and CAO conditions, I claim, are both necessary for cognition of any (or almost any) type. I suspect, moreover, that the CAO condition as I have cursorily stated it here could be refined so as to make the GAA and CAO conditions jointly sufficient for cognitive activity, though I do not take up the task of providing the needed modification here.[8]

IV. COMPUTATIONALIST CONCLUSIONS

It is by now clear that I believe there is a strong analogy between Olympia and a dreamer/bell system, and thus an interesting case to be made against Maudlin's claim that Olympia actually performs a computation on the input τ. Olympia, as noted above, undoubtedly meets the GAA condition for computation, but she does not meet the CAO condition for computation. The input τ does not function as an active cause of Olympia's W/M/CS activity when this input is given, hence there is no reason, it seems to me, to count this activity as a computation performed on the input τ. Olympia is engaged in an activity that is too "introverted," in a sense, to amount to a computation; since the active causes of her W/M/CS activity are supplied by Olympia's own cognitive machinery, her activity amounts to the machine analogue of a dream, or hallucination, or act of the imagination. Computation is, like virtually all cognitive activities, an activity that supervenes on a particular sort of causal interaction between a cognizing subject (here, the computational machinery) and a cognized object (here, an input given on a tape). Computation, moreover, is intrinsically a reciprocal, active causal interaction in that it requires that the machine actively transform the tape's contents (by writing operations) and that the tape's contents actively work to "transform" the machine's activity (its movements along the tape, its writing operations, and its state transitions) via the tape reading process. Olympia engages in the first, but not the second, form of active causal activity. But it is, again, only in virtue of the fact that the object actively induces the subject's activity that it is correct to describe the activity as a cognition (here, a computation) of the object. There is, in conclusion, no fact about Olympia's W/M/CS activity, or her architecture, that establishes her activity on τ as a computation performed on the input τ, despite the fact that this W/M/CS activity is the same activity that Olympia would have engaged in had she actually performed such a computation.

---

[8] It seems to me that what the CAO condition ought to say is not simply that the cognized object function as an active cause of the cognizing activity, but that the object do this in some particular way. I leave aside the problem of characterizing this way.

For all that has been said, perhaps one's critique of a computationalist theory of consciousness might resiliently insist that the GAA condition alone is sufficient for cognition, and thereby concede that, in the dreamer/bell/text scenario, the dreamer/bell system is, after all, reading the text (and, in the dreamer/bell/candle scenario, the dreamer/bell system actually refers to the real candle). My only response to this reply is that, for me, reflection on the dreamer/bell scenarios is sufficient to make such a proclamation sound somewhat dogmatic. I would be happy to avoid dogmatism altogether and simply speak of GAA cognition (which asserts that the GAA condition is sufficient for a cognitive act) as well as GAA/CAO cognition (which takes the GAA condition and the CAO condition as both necessary and sufficient under some suitable modification of the CAO condition, for cognition). Perhaps both senses of cognition are legitimate (though I will confess I doubt it).

Olympia, while she succeeds in performing a GAA computation, does not succeed in performing a GAA/CAO computation. Hence Maudlin's argument entails the following conclusion: the status of a system as conscious cannot supervene on a system simply in virtue of its GAA computational structure. There is nothing in his argument which counts against the possibility that a system is conscious simply in virtue of its GAA/CAO computational structure. For those of us who take 'GAA/CAO computation' as simply equivalent to the ordinary notion of "computation," this amounts to Maudlin's failure to refute a computationalist theory of consciousness.

ERIC BARNES

Denison University

BOOK REVIEWS

Innocence and Experience. STUART HAMPSHIRE. Cambridge: Harvard University Press, 1989. 195 p. Cloth $20.00.

Moral philosophy, according to Stuart Hampshire, takes "prevailing moral ideas" as the "raw data and starting-point" for a critical reconstruction of "those few well-argued conceptions of human powers which now actually influence people's lives and purposes" (179, 32). Hampshire draws his prevailing moral ideas from literature and history as well as from an extraordinarily wide-ranging and humane moral reflection. He distills from these ideas an account of human nature from which he draws normative conclusions about "the perennial topics of the underpinnings and origins of justice, of the universal and conventional elements in justice, and of the relation of private to public morality" (157). *Innocence and Experience* recapitulates several themes and arguments from Hampshire's earlier books, but it also contains much that is fresh. Although not all its arguments succeed, the book's thoughtfulness and intellectual vigor are remarkable, and it offers a systematic statement of what should be regarded as a leading position in moral and political philosophy.

I

The essence of humanity, Hampshire argues, is "a perpetual openness to new ways of life, to new forms of thought, to innovations in language and social arrangements" (30). This capacity for "imaginative invention" inevitably drives us to formulate diverse and irreconcilable conceptions of the good. In Hampshire's view, imagination ranks alongside intellect, which "unites humanity across all barriers" (30) just as imagination divides us. He emphatically rejects all topologies of the soul that place reason at the top of a hierarchy of faculties. The equal primacy of imagination and intellect as defining features of the human condition implies a "dual aspect" theory of morality (136). The first aspect is pluralism and sempiternal conflict among substantive visions of the good, since these are the products of imaginative invention and the centrifugal forces of linguistic diversity (43) and memory. (Hampshire argues that our memories are the source of "singularity" (114–5) and ultimately of imagination (127).) Joined with plural visions of the good, however, is a thin (137), universal (77, 140), transhistorical (142), species-wide (154) notion of minimum procedural justice consisting in the willingness to submit to "a regular and reasonable procedure of weighing

0022-362X/91/8806/317-24

claims and counter-claims" (63) through negotiation or arbitration or adjudication. Thus, Hampshire's prescription is this: "every population of intelligent persons should be a patchwork of minorities in the pursuit of deviant conceptions of the good, held together by a shared respect for fairness" (132). Strikingly, he concludes in favor of "Heraclitus's vision" that "ineliminable conflicts . . . and . . . rationally controlled hostilities . . . [are] not only normal, but also the best condition of mankind from the moral point of view" (189).

Just as pluralism of the good is a consequence of human imagination and diversity, minimum procedural justice "is rooted in" (169) the "species-wide and cross-cultural endowment" (119) of practical rationality within a single individual's deliberations (53, 169, 181, 187). Practical rationality Hampshire describes as "the practice of promoting and accepting arguments for and against a proposal" (53). Like Plato, Hampshire regards "governance of the soul and governance of the state" as mirrors of each other (182).

This is a dubious identification, however, for it overlooks a decisive disanalogy between the ways individuals and groups weigh arguments. Arguments in negotiation and adjudication are invariably strategic: the contending parties argue in order to prevail, not in order to achieve the fairest, best, or most informed outcome. In adjudication, the litigants employ rules of procedure tactically to exclude or conceal inconvenient evidence; in negotiation, the parties hide the weaknesses of their positions and disguise their bottom lines. Thomas Schelling[1] has pointed out that a bargain can be struck only when one party concedes, and a party concedes only because she believes that the other party will not. It follows that no bargain can be struck if the parties know what the others will actually accept; agreement requires bluffing. Bargaining and adjudication, even in good faith, are defined by opacity and cunning, not openness. As Hampshire characterizes it, practical reason is dialectic, in the sense of Socrates or Karl Popper: "a precise analogue of the requirement in a theoretical inquiry that all the accessible evidence should be assembled and assessed" (94). In adjudication, arbitration, and negotiation, by contrast, argument is rhetorical. While it may also be true, and interestingly so, that an individual deliberating about practical decisions often deceives herself and makes rhetorical appeals *in foro interno*—Hampshire suggests this in the book's final paragraph—the contending standpoints within a single person are nonetheless controlled by her overriding interest in making the best decision. This is almost never true in adjudication or negotiation.

---

[1] *The Strategy of Conflict* (Cambridge: Harvard, rev. ed. 1980), pp. 21–2.

Moreover, adjudication and negotiation are in important respects inconsistent with the openness to all claims which Hampshire promotes. The procedural rules of adjudication grant jurisdiction over some matters but (therefore) foreclose jurisdiction over others; the most important decisions of courts are often exercises of the "passive virtues," the studied refusal to hear issues on the merits because the court believes that a definitive ruling would be bad for the political process.[2] Similarly, the possibility of negotiation typically turns on excluding parties and issues from the table. Gag rules make procedural justice possible.[3]

Hampshire's derivation of minimum procedural justice from the essential features of human nature thus fails; whatever minimum procedural justice is "rooted in," it is not individual practical reasonableness. One should not assume, however, that this seriously damages the case for minimum procedural justice. Openness to argument and willingness to entertain seriously the claims and interests of others, including enemies and foreigners, are clearly appropriate candidates for universal moral requirements, particularly since violence and oppression seems to be the sole alternative. In Churchill's words, "Jaw-jaw is better than war-war." Nor is there anything minimal about minimum procedural justice: listening in good faith to the arguments of all interested parties is an ideal that no historical government has even remotely approached.

It is hard to see why procedural justice is the *only* universal requirement, however, as Hampshire himself comes close to acknowledging when he defines "great evils" as "those states of affairs which are to be avoided for reasons that are independent . . . of any specific conception of evil" (106). These include "[p]hysical suffering, starvation, imprisonment, the destruction of one's family or home," and he argues that these "have a clear priority in practical thinking" (106). This being so, it is somewhat puzzling that Hampshire stops short of the conclusion (pressed by Bernard Gert in *The Moral Rules* and *Morality*) that the requirement to refrain from inflicting great evils constitutes a minimum and universal *substantive* morality. One might object to this requirement that the infliction of the great evils may in some circumstances be justified; but this cannot be Hampshire's objection, for he insists that the violation of minimum *procedural* justice may in some circumstances be justified as well (140, 168, 171–2, 186). Universal duties are

---

[2] See Alexander Bickel, *The Least Dangerous Branch: The Supreme Court at the Bar of Politics* (Indianapolis: Bobbs-Merrill, 1962), pp. 111–98.

[3] See Stephen Holmes, "Gag Rules or the Politics of Omission," in Jon Elster and Rune Slagstad, eds., *Constitutionalism and Democracy* (New York: Cambridge, 1988), pp 19–58.

"absolute," but only in the sense that they are not relativized to particular theories of the good, not in the sense that they can never be overridden (140). Moreover, the familiar fact that procedural rules inevitably affect, and often determine, substantive outcomes partially undermines the reliance Hampshire places upon the distinction between procedural and substantive justice.

His explication of minimum procedural justice is a bit perfunctory. He neglects the important differences between adjudication and arbitration, on the one hand, and negotiation, on the other. In Hampshire's own terms, adjudication—which yields all-or-nothing resolutions, winners and losers—is the procedural justice of innocence, whereas negotiation—always aimed at compromise and log rolling—is the procedural justice of experience. More troubling is his neglect of the culturally-relative character of procedural justice, the fact that adjudicatory systems proliferate in much the same way as theories of the good. (Indeed, on Hampshire's argument, they would have to. The machinery of procedural justice itself "has to be established by negotiation" (109), and thus must bootstrap itself into existence. This implies that systems of procedural justice are imaginative inventions, not intellectual deductions, and will therefore assume diverse forms.) The plurality of procedural justice systems weakens Hampshire's case for bifurcating pluralistic substantive morality from universal procedural justice.

II

Throughout the book, but principally in chapter four, Hampshire offers a highly original and compelling *catalogue raisonné* of human passions and powers. His discussion of the vast yet neglected role that reminiscence and nostalgia play in value formation is superb. Hampshire's scope is wide, and he has evidently written at a high level of inspiration; in the course of demarcating the places of imagination and intellect in human life he offers remarkable insights about such disparate topics as sexual love, language acquisition, nationalism, Nazi anti-Semitism, and philosophies of history. In breadth and depth, *Innocence and Experience* is reminiscent of Hannah Arendt's *The Human Condition* (without the Hellenism) and Roberto Unger's *Passion* (without the romanticism). All three books engage in the same project of drawing normative conclusions from a broad-gauge account of human nature—what Unger calls "the classical style of normative argument"—and it is noteworthy that they concur in the decisive role they assign to human inventiveness and openness.

The classical style of normative argument rests on the possibility of arguing from facts to values, and Hampshire devotes a large part of the book to vindicating this procedure by exorcising "Hume's ghost," which insists that the entire enterprise rests on a mistake. He responds that Hume erroneously conceives of moral evaluations as "commentary on, and criticism of, human performances" (102) rather than as steps in practical deliberation. This "displacement of agency into observation" (18) leads Hume to "the error of false isolation" (82) of value judgments from the contexts that alone give them their sense (83–8), and it is this error that leads Hume to deny the rationality of value judgments.

Hampshire supports these claims with two more detailed arguments. First, he advances a metaethical argument that, in its attributive uses, the word 'good' is not the name of a property but rather the adumbration of a "conditional wanting formula" (24) according to which "*x* is a good *S*" means roughly "if you want an *S*, *x* is what you are looking for." Thus, there is a systematic connection between the description of *x* as an *S* and the evaluation of *x* (95). Second, Hampshire returns to an argument he has made elsewhere, stressing that in practical reasoning we typically deliberate about singular statements of possibility: we wonder whether *X* would happen if we *Y*, or whether it would be morally wrong to *Z* (where *X*, *Y*, and *Z* are unique, nonrepeatable events). The former is "predictive" and the latter "moral," but Hampshire argues that the former is no more objective than the latter.[4]

It is unclear that Hampshire has succeeded in bridging the fact-value gap, however. Although there may be a "systematic connection" (95) between "*x* is an *S*" and "*x* is a good *S*," there is none between "*x* is a good *S*" and "*x* is (morally) good." (Let '*S*' equal 'Nazi'.) Thus, the first argument eliminates the gap between "*x* is an *S*" and "*x* is a good *S*" only to relocate it between "good" in its extramoral and moral uses. And the second argument eliminates the difference in objectivity between (singular subjunctive) predictions and (singular subjunctive) evaluations only to relocate it between singular subjunctives of either sort and scientific laws. In the end, this much of the gap remains: we always face the choice between accommodating human nature and resisting it, and morality can be read off of human nature only on the accommodation strategy. Aristotle's is an accommodation strategy, whereas ascetic religions adopt resistance strategies. Hampshire's conditional wanting analysis of "good," which rules out the claims (made by skeptics but also by mystics) that what is good may be wholly antithetical to human wants, implies an accommodationist commitment that cannot itself be derived from the facts of human nature without begging the question.

---

[4] Hampshire first ventured this argument in "Subjunctive Conditionals," *Analysis*, X (1950): 9–14, repr. in *Freedom of Mind* (Princeton: University Press, 1971).

III

Along with the key distinctions he draws between imagination and intellect and between thickly substantive and thinly procedural moralities, Hampshire is passionately concerned with the moral contrast between innocence and experience. "The idea of experience," Hampshire writes, "is the idea of guilty knowledge, of the expectation of unavoidable squalor and imperfection, of necessary disappointments and mixed results, of half success and half failure" (170). There is no question that Hampshire adopts the standpoint of experience, but his admiration for the virtues of innocence is genuine (172–7), and he insists that the standpoint of innocence—interestingly characterized as freedom "from the strains of a perpetual contrivance of means to ends" (149)—is the source of values essential to a tolerable life. Indeed, Hampshire speculates that nostalgia for lost innocence may be identical with our idea of pure happiness (149). Perhaps the most interesting and important moral problem Hampshire addresses (and it is noteworthy that few other contemporary philosophers even raise it) is how we can confront the truths of experience without falling into disillusion or cynicism.

These reflections inevitably direct Hampshire to the problem of dirty hands in politics. Hampshire, like Machiavelli, has little patience with the voice of innocence that requires politicians to abjure "deceit and guile, unjust violence and sudden aggression, ingratitude in relations with allies and friends" (163). Machiavelli argued that a prince cannot live up to Christian morality; Hampshire makes the problem harder still by asking whether the prince can respect even "the bleak negations of procedural justice" (164). His answer is yes, even though he insists that

> [a] political leader who allows a commitment to fair dealing and to substantial justice in all circumstances to override his determination to retain power, and to protect the interests of his country, will surely become ineffective and impotent; and such a person would generally have done better never to have sought or accepted public responsibilities (164–5).

His argument contains two points. First, "evil, in the form of the drive to domination, consists in the *uncompensated* violation of this basic justice" (186, emphasis added). But sometimes violations of procedural or substantive justice are "compensated" because the alternatives are worse. (And thus the violations are not evil? But Hampshire also refers to such decisions as striking a balance between evils (171). Thus he does not seem to have fully faced up to the problem of dirty hands, the paradox arising from the claim that in politics it may be right to do wrong.) Hampshire takes as his example the unjust victimization of a minority group in order to maintain the standard of living of a leader's constituents (186; also 171). Hampshire is prepared to concede that this injustice may under some circumstances be justified, for politicians have "obligations to protect the reasonable interests of their innocent fellow citizens" (170) even by ruthless means. This he calls the "Machiavellian thesis."

His second point is that determining the lesser evil requires the politician to weigh all the contending arguments, so that "the decision is not made by default" (171). Thus, in the weighing of evils and of interests, minimum procedural justice "plays the role of the scales" (186). More specifically, the leader must "establish the nature of the claims upon him in their own terms before he restates them in his own terms drawn from his own conception of the good" (187). Even ruthless and unjust political behavior must respect minimum procedural justice by hearing out the arguments.

It appears, however, that Hampshire has made the problem too easy, too "innocent," in two distinct ways. First, in real life, political expediency may well require the politician to violate not only substantive justice but even minimum procedural justice in order to govern effectively. (Hampshire himself seems to recognize this when he asserts (164–5) that expediency may be inconsistent not only with substantial justice but also with fair dealing.) Floor managers at political conventions labor to deny dissidents the opportunity to present their case; presidents appease their coalition members by refusing to meet with the members' adversaries; diplomats maneuver to shut inconvenient groups out of negotiations. All of them claim that these violations of minimum procedural justice are necessary for political success, and we have no reason to doubt them.

Second, the "Machiavellian thesis" speaks of obligations to protect constituents' interests, and seems to presume that only obligations of this sort will justify injustices. Thus, Hampshire's problem arises only when the politician "weighs two moral claims on him" (186). But the practicing politician confronts a meaner problem: he or she must commit injustices not only in order to fulfill moral obligations, but simply in order to acquire and retain power. Nor need the acquisition and retention of power be morally motivated; many politicians have no large ambitions toward the public good, but merely expect that they will muddle through their term in office as creditably as their opponents would. Thus, the *real* Machiavellian problem may be this: What moral sense are we to make of a world in which mediocre politicians must violate minimum substantive and even procedural justice simply to hang on to their jobs?

Despite these problems, *Innocence and Experience* is a wise and generous book that it would be difficult to praise too highly. Calmly and beautifully written, it manages to say important things about a wide range of human experiences within a very few pages. Notwithstanding his Blakean title, Hampshire is neither visionary nor romantic; his book seems more akin in spirit to Shakespeare, with whom he shares an affection for the frailties as well as the victories of the life we lead in a decidedly nonutopian world. *Innocence and Experience* may well fall into the select circle of books whose admirers will find their conversations beginning with the question, "Which are *your* favorite parts?" (For the record, mine are the sections "Nazism and Evil," "Attachment to the Past," and "Innocence and Experience Illustrated.")

DAVID LUBAN

University of Maryland

Metaphor: Its Cognitive Force and Linguistic Structure. EVA KITTAY. New York: Oxford University Press, 1987. 358 p. Cloth $59.00.

Eva Kittay argues that metaphors have a "second-order" metaphorical meaning in addition to their literal meaning, and that it is possible to provide a theory of interpretation for this second-order meaning. Thus, she argues against approaches such as Donald Davidson's,[1] in which it is argued that metaphors *mean* what they literally express, but that they may be *used* (for example) to call us to compare two objects or events.

Kittay's specific proposal is that the literal meaning of an utterance is to be fixed by the role that the utterance plays in some contextually determined *semantic field*—where a semantic field consists in the dependencies holding between concepts in a particu-

---

[1] "What Metaphors Mean," in *Inquiries into Truth and Interpretation* (New York: Oxford, 1984), pp. 245–64.

0022-362X/91/8806/324–30

lar domain. Metaphoric meaning (or second-order meaning) then arises from the introduction of relations from one domain or semantic field (the domain of the vehicle) onto another semantic field (the domain of the topic). The idea can be thought of as an elaboration of a suggestion due to Max Black[2] that metaphors introduce new relations into a "system of commonplaces."

My comments are divided into two sections. In section (I), I consider Kittay's specific proposal—in particular her thesis that metaphor consists in the transfer of relations from one semantic field to another. In section (II), I return to the more general issue of whether metaphor is a matter of use or interpretation, considering a number of arguments that Kittay offers in response to Davidson's "What Metaphors Mean."

I

To set up the discussion, an example of how Kittay applies her theory is in order. In Plato's analogy of Socrates as midwife, the semantic field of child delivery (according to Kittay) is structured as follows:[3]

- main verb: help
- agent: midwife
- resulting state of affairs:
  - verb: create (deliver)
  - agent: mother
  - result: child
- instrument: potions and incantations

This structure is then imposed onto the semantic field of philosophical instruction in the following way:

- main verb: help
- agent: Socrates
- resulting state of affairs:
  - verb: create (deliver)
  - agent: student
  - result. ideas
- instrument: dialectic

This is an interesting way to think about metaphor, and it gives some substance to Black's idea of transferring relations from one system of commonplaces onto another. The transferred relations turn out to be the grammatical relations of agent, patient, instrument, etc., which linguists call *thematic roles*.[4]

---

[2] "Metaphor," in Black, ed., *Models and Metaphors* (Ithaca: Cornell, 1962). Kittay, I should note, also takes issue with Black on a number of points; see 31, 181.

[3] This sort of analysis is also found in her joint work with Adrienne Lehrer. See Kittay and Lehrer, "Semantic Fields and the Structure of Metaphor," *Studies in Language*, v (1981): 31–63.

One important concern, however, is whether the above analysis is interesting because of semantic fields or because of the use of concepts like agent, patient, and instrument. Strictly speaking, these concepts may not be incompatible with the theory of semantic fields, but you do not need to believe in semantic fields to employ such concepts. If thematic roles are doing all the work, why do we need the semantic fields? In chapter 6 other kinds of semantic fields are introduced—in particular, contrast sets (e.g., the following is a contrast set, ⟨verbs of cooking: boiling, baking, frying, steaming, etc.⟩) —but thematic structure does most of the work in chapter 7 when semantic fields are applied to metaphor.

Kittay often speaks of the advantages of semantic fields and their holism over "atomistic" semantics, but cannot help making use of the resources of atomistic semantic theories at critical junctures. For example, in chapter 2 (59) she utilizes the notion of selectional restrictions from Katz-Fodor[5] semantics—which would have to be considered a paradigm example of an atomistic semantic theory. Elsewhere (58–64), she assigns interpretations to utterances *via* what looks like a compositional semantics.

This point is made more significant by the fact that semantic fields are given no real independent justification in the book. Kittay (35, 216), like a number of continental philosophers, uses an analogy to phonology and cites Ferdinand Saussure's "insight" that there are no phonetic primitives and that phonemes are defined by the role that they play in a system of oppositions. Strictly speaking, this is not an insight, but an empirical hypothesis, and the hypothesis has been in eclipse since the development of the theory of distinctive features by Roman Jakobson in the 1940s. A central point of modern generative phonology has been that there is a stock of primitive features common to all human language from which the phonology of a given language is derived. Modern phonology is in fact atomistic.[6]

---

[4] The theory of thematic roles has its roots in Charles Fillmore's case theory, see "The Case for Case," in E. Bach and R. T. Harms, eds., *Universals of Linguistic Theory* (New York: Holt, Rinehart, & Winston, 1968), pp. 1–90.

[5] See Jerry Katz and Jerry Fodor, "The Structure of a Semantic Theory," *Language*, xxxix (1963): 170–210

[6] Stephen Anderson [*Phonology in the Twentieth Century* (Chicago: University Press, 1985), pp. 118–9] puts Jakobson's perspective as follows:

> . Jakobson felt that the doctrine of the arbitrariness of the sign had to be subjected to important qualifications. While it is of course true that signs are arbitrary in that the link between a particular *signifié* and a particular *signifiant* is established by the conventions of a particular linguistic system, this

Independently of whether semantic fields are necessary, there are other concerns about the specific proposal. One such concern may be found in the following objection from Jonathan Adler. Suppose I am trying with difficulty to cut my steak and say, "this knife is a spoon." Either Kittay must claim that my utterance is not a metaphor, which seems false, or she must claim that I have introduced relations from one semantic field onto another. The problem with the latter solution is that knives and spoons intuitively belong to the same semantic field. If they do not belong to the same field in this case, then we may rightly ask for criteria of individuation for semantic fields. The danger of circularity should be obvious here. Kittay must avoid using the presence of metaphor as a criterion for individuating two distinct semantic fields in cases like the above.

There is perhaps another solution to this difficulty. There is really no need to say that the new relations have to come from *another* semantic field. One could very well say that the relations merely come from some semantic field or other (possibly the same one). In the knife and spoon case, it is possible to say that one is imposing the relations of a semantic field back onto itself, with the change being that the objects in the field are taken to participate in different relations.

Although not specifically related to the use of semantic fields, it is worth noting that Kittay takes the rather courageous view (64 ff.) that metaphor essentially involves some sort of semantic anomaly. The view is courageous, for it must account for apparently non-anomalous metaphors like 'Liddy is Nixon's plumber'. Kittay argues that the semantic anomaly of these constructions will become appar-

---

> does not imply that the sorts of things which can serve as potential *signifiants* are constrained only by the necessity to be different from one another. In particular, language has an essentially spoken character, as a result of which the *signifiants* are necessarily to be found in their logically primary form in the structuring of sound. Furthermore, not just any differences in sound are potentially relevant phonologically: a universal inventory can be given of the small set of distinctive features that function to differentiate *signifiants* in natural languages. Insofar as the *signifiant* of any sign must be made up of these same elements, the arbitrariness of the sign is considerably restricted.

Important work echoing the same perspective includes Roman Jakobson, C. Gunnar, M. Fant, and Morris Halle, ***Preliminaries to Speech Analysis: The Distinctive Features and their Correlates,*** MIT Acoustics Laboratory, Technical Report No. 13, 1952, repr. Cambridge: MIT, 1963; Noam Chomsky and Morris Halle, ***The Sound Pattern of English*** (New York: Harper & Row, 1968); and Halle, "On Distinctive Features and their Articulatory Implementation," ***Natural Language and Linguistic Theory,*** I (1983): 91–105. I am indebted to Ellen Broselow, Richard Larson, and Dan Finer for discussion of these matters.

ent in a fuller discourse—one in which we are told, for example, that Liddy is a lawyer. It is not clear that this move will succeed, however, for it seems to collapse the distinction between falsehood and semantic anomaly.

II

Kittay devotes a fair portion of her book to an attack on Davidson's theory of metaphor and his thesis that metaphor is part of the domain of language use, rather than meaning. Most of her criticisms seem to assume that Davidson's arguments are inextricably tied to his program in semantics, but he in fact offers several arguments in "What Metaphors Mean," which are independent of that program. We can call them the ***argument from simile,*** the ***argument from dead metaphor,*** and the ***full paraphrase dilemma.***

The argument from simile goes roughly as follows (*op. cit.*, pp. 253–7). We are seldom tempted to invoke figurative or second meanings to account for similes, for it is clear in such cases that the author of the simile is inviting us to make a comparison between two things—and is not exhaustively specifying the properties shared by those things. Why should we not take the same stance toward metaphors? Kittay notes this argument, and her response is as follows.

> Simile accomplishes some of the same cognitive ends [as metaphor] but uses different linguistic resources. A simile is not a connotative semiotic and has no second-order meaning. But the nature of the comparison is metaphorical. In particular, the 'like' in simile should be understood metaphorically, for it signals a *cross-categorical* comparison (143, fn.).

Of course, this passage simply hands the point to Davidson, for it concedes that no second-order meaning is necessary to make metaphorical comparisons.

The argument from dead metaphor is simple enough. Davidson argues that, if there really is a special (let us say second-order) meaning, why does that meaning not become the literal meaning when the metaphor dies (*op. cit.*, pp. 253, 261)? For example, the phrase 'mouth of a bottle' presumably began as a metaphor, but whatever metaphorical meaning it once had is surely not to be found in the dead metaphor. Kittay, takes note of Davidson's argument, but seems to miss the point.

> Davidson suggests that if metaphors had a special meaning then 'we might be able to specify the special meaning of a word in a metaphorical setting by waiting until the metaphor die[d]' [Davidson, 1984, p. 253]. While it is true, as Davidson says, that this cannot be done, we ought not thence to conclude that metaphors have no special meaning. Rather, if instead of a second meaning, as in the case of ambiguity, they

> have a second-order meaning, we understand why this cannot be done. For a genuinely dead and unrevivable metaphor is one in which no second-order meaning is recognized by speakers of the language—the word acquires as a first-order meaning the sense which previously only contributed to its second-order meaning (142).

Davidson's point is that, on theories like Kittay's, the first-order meaning of the dead metaphor *should* be the same as the original second-order meaning, but of course it is not (as the mouth of the bottle example among others shows). And even if metaphors never die but merely lie dormant, the question arises as to why they never take on the second-order meaning while dormant.[7]

The full paraphrase dilemma is arguably the central point of Davidson's paper.

> There is, then, a tension in the usual view of metaphor. For on the one hand, the usual view wants to hold that a metaphor does something no plain prose can possibly do and, on the other hand, it wants to explain what a metaphor does by appealing to cognitive content—just the sort of thing plain prose is designed to express. As long as we are in this frame of mind, we must harbour the suspicion that it *can* be done, at least up to a point (*op. cit.*, p. 261).

Kittay must either admit that metaphor can do no more than plain prose, or give up the notion that there is a definitive cognitive content (i.e., a second-order meaning) attached to metaphor. She maintains that metaphors do not admit of paraphrase (thus rejecting the first horn of the dilemma), and argues that they cannot be paraphrased because of the nature of semantic fields.

---

[7] James Higginbotham has suggested that Davidson's argument assumes too much about what it means for a metaphor to die. Suppose we reconstruct Davidson's argument as follows.

> Suppose an expression E has a literal meaning M and is used metaphorically. By hypothesis, the metaphoric meaning is M'. Furthermore, if E "dies" we would expect M' to be its meaning. But when E dies it does not have the meaning M' but rather M'' whose reference is that of M'. Conclusion: E never had the metaphoric meaning M'.

In the example above:

M: (literally) the mouth of the bottle
M': what stands to a bottle as the mouth does to an animal
M'': the opening in the top of a bottle

The defender of metaphoric meaning might hold that there is no other way to get at M'' other than through M'. Thus, the burden is shifted back onto Davidson to show how one gets at the referent of the metaphor without assuming the metaphoric meaning. I think Davidson can uphold this burden. Just as it is often argued that names and demonstratives need not refer via descriptions, it can be argued that dead metaphors refer to something without the help of metaphorical meaning to fix the referent.

> We can give an exposition of a metaphor but not a paraphrase—not if the metaphor is still a live metaphor and has not collapsed into literality. The irreducibility of metaphor is importantly tied to the incongruity between the domains of the topic and vehicle. That incongruity guarantees that a metaphorical predication cannot easily accommodate itself in the conceptual scheme which lies behind literal and conventional language (37).

One might ask why incongruity between two semantic fields "guarantees" that the relations between them cannot be characterized in literal language.[8] We are given no argument. But if Kittay is correct about this—that the incongruity of semantic fields prevents paraphrase—then she has new worries. Kittay is, after all, presenting a theory of metaphors, and is describing specific cases of metaphors in a book that is largely composed of literal prose. If there is more to a specific metaphor than she can describe in the language of her theory, that is just to say that her theory of second-order interpretation does not completely explain what the metaphor does. Moreover, to the extent that any theory of metaphoric interpretation will have to be couched in literal language, it will follow that there can be no theory that explains what the metaphor does. But this is just Davidson's point—that no theory of interpretation can explain what a metaphor does.

III

There are a number of interesting ideas in this book, and if the arguments do not always succeed, they at least advance the discussion. A number of the ideas can be divorced from the use of semantic fields, which, appearances aside, are not essential to Kittay's project for metaphor. If there is a problem with the book, it is that it underestimates the strength of Davidson's position. Kittay is not alone in this. Many philosophers working on metaphor are guilty of the same.

PETER LUDLOW

State University of New York/Stony Brook

---

[8] Kittay does not say metaphorical predication *cannot* be accommodated in literal language but merely that it cannot *easily* be accommodated in literal language. But if it is merely difficult to accommodate these predications in literal language, it would seem that one in fact *can* provide a paraphrase of the cognitive content of metaphors, or come as close as one likes.

Hilbert's Program: An Essay on Mathematical Instrumentalism. MICHAEL DETLEFSEN. Boston: Reidel, 1986. xiv, 186 p. Cloth $42.00.*

This clearly written book has two main goals: (a) to set forth an "Hilbertian" epistemology of mathematics, i.e., a theory of mathematical knowledge extracted from the writings of David Hilbert; (b) to show, contrary to the received wisdom, that Kurt Gödel's theorems have not discredited this mathematical epistemology.

Michael Detlefsen, through Hilbert, does to realism in mathematics what Bas Van Fraassen recently did to scientific realism: namely, just as Van Fraassen argues that scientific theories need not be, and anyhow cannot be shown to be, anything more than "empirically adequate," so does "Hilbert" argue that mathematics, to be good, need be only "real sound"—i.e., true to the so called "real" propositions of mathematics. The "real" propositions, are, roughly, arithmetic and algebraic (quantifier-free) identities. Since some propositions for Hilbert are, in fact, real, he occupies a middle ground between fictionalists like Hartry Field and realists like Gottlob Frege.

"Ideal" propositions and "ideal" proofs, corresponding to "theory" in Van Fraassen's scheme, are to be treated instrumentally: evaluated (as epistemic tools) rather than interpreted. This evaluation is done, for "Hilbert," in a theory called "metamathematics," which is itself supposed to be free of "ideal" proofs and propositions to prevent what Detlefsen calls (cf. 16ff., 62ff.) the epistemic "dilution" of mathematics. (The less squeamish Van Fraassen does his evaluations in model theory, which is applied set theory.) In fact, Hilbert suggested, a metamathematical consistency proof would be, not only necessary, but also sufficient, for the evaluation. His reasoning was: if mathematics is not "real sound" then it can "prove" a false identity; but any such false quantifier-free formula has, as a counterexample, a (provable, of course) true arithmetic identity. Thus, if mathematics is not real sound it is inconsistent. Similarly, a consistency proof allows one to replace any ideal proof of a *real* proposition *P* with a *real* proof of *P*: if *P* were false then mathematics would be inconsistent—contradicting the "real" consistency theorem. Thus, a consistency proof for Hilbert plays a role some-

---

* I am indebted to Haim Gaifman, Saul Kripke, Sharon Shelah, Charles Parsons, and Stewart Shapiro for extremely valuable comments and criticisms of my earlier versions.

0022-362X/91/8805/331–6

what like the existence of God in Descartes's *Meditations:* a guarantor of proved propositions.[1]

Now, just as Descartes did not intend to replace science by theology, Hilbert does not claim that we can or should replace mathematics by metamathematics. Rather, mathematicians continue doing what they have always done: discovering new theorems. They could hardly make these discoveries in metamathematics. The "ideal" formalisms of mathematics have other virtues than that of being, say, true—they are instruments suited to human inquiry.

As the Yiddish saying goes: if things are so good, why are they so bad? Why is it that the Hilbert program is considered "dead" or, at best, replaced by other research programs? Gödel's theorems, of course, especially the second. This purports to show that one cannot prove the consistency of any past or future actual mathematical theory within formal number theory. Presumably, formal number theory contains all the "real" proofs there are in mathematics.

Detlefsen, in a careful epistemological analysis, spells out this "standard strategy" for defeating Hilbert's program. I shall merely say that both the exposition of his position, and the analysis of the "Gödelian objection" to it are so clearly written that even the beginner in philosophy of mathematics could benefit from reading the book.

He has two methods of defeating the received wisdom, within the spirit of what he calls "Hilbertian instrumentalism." One is to note that, if "ideal" mathematics is to be taken seriously as instrument, then it must be suited to the finite capacities of human beings. The human brain cannot make use of all of the potential proofs in "ideal" mathematics. It is at least possible, then, that all the "Hilbertian" need do is prove the consistency of some subclass of all the possible proofs in mathematics—and there might well be a metamathematical proof that this subclass is consistent. This is the argument Detlefsen develops in chapter V. As Detlefsen points out, this is no argument for Hilbert, but only a refutation of the "Gödelian challenge." I think, moreover, that it would be difficult to designate the subclass of "ideal" proofs to be proved consistent, because we can always introduce new concepts and thus proofs by abbreviations, as is done in informal mathematics. Formally, the use of abbreviations makes it possible to know the existence of "ideal" proofs other than by actually exhibiting the proof in the basic language of the system.

Detlefsen's other "instrumentalistic" rejoinder to Gödel is to deny the received opinion that Gödel's second theorem demonstrates the impossibility of proving the consistency of "ideal" mathematics in "real" mathematics. Here his argument, as set forth in chapter IV, takes two prongs:

---

[1] A major difference between Hilbert and Descartes, though, is that Descartes regards the absolutely infinite as clear, distinct, and real; Hilbert admits as "real" only propositions concerning the potentially infinite.

(a) Gödel shows that a certain number-theoretic sentence, which can be "interpreted" as the claim that "mathematics is consistent," cannot be proved—in number theory, and hence in "real" mathematics. He does not show that *every* sentence that expresses the consistency of mathematics is similarly unprovable. And although a number of writers have tried to generalize Gödel's theorem to show that every plausible candidate for the honor of expressing the consistency of mathematics is Gödel-vulnerable, Detlefsen, examining their arguments one by one, concludes that they fail to eliminate all possible candidates.

Detlefsen focuses his attention on the so-called "derivability conditions" which have figured in discussion of Gödel's theorems. Consistency of (a given system of) mathematics is, of course, the statement that $0 = 1$ is not provable (in the system). Suppose we want to represent the notion of "provability" (of some suitable system) by a formula of arithmetic, call it Provable(x). Now it turns out that Gödel's theorems apply to any formula that obeys the following three conditions:[2]

I) If X is a theorem of mathematics, and [X] is (the numeral denoting) its Gödel number, then Provable([X]) is a theorem of arithmetic.

II) For every formula X of mathematics, it is a theorem of arithmetic that

If Provable([X]) then Provable([Provable([X])])

This amounts to saying that condition I is itself formalizable in arithmetic.

III) For formulas X and Y of mathematics the following is provable in arithmetic:

If Provable([X]) and Provable([$X \rightarrow Y$]) then Provable([Y])

This amounts to saying that the principle of modus ponens is formalizable in arithmetic.[3]

Mostowski and other writers have argued that any arithmetic formula that does not obey these three conditions cannot formalize the intuitive notion of mathematical provability. Since these three conditions are also sufficient to produce Gödel's theorem, their conclusion is that Gödel's theorem refutes Hilbert's program. Detlefsen argues, plausibly in my view, that no one has yet given a convincing reason that precisely these three conditions express the "essence" of provability. He therefore concludes that the case against Hilbert from Gödel's theorem falls short of demonstration.

---

[2] I am omitting discussion of the so-called "Diagonalization" property, which seems to me a property of the system, not of provability.

[3] These principles bear a formal resemblance to principles of modal logic, a resemblance noted by Kripke, Georg Boolos, and others, but I shall not discuss the significance of this analogy.

I think that most logicians will be unpersuaded by Detlefsen here (by 'most logicians' I mean Haim Gaifman and Saul Kripke); for I do not think that the derivability conditions are really the heart of the matter.

The reason that derivability conditions are needed to winnow out candidates is that, in arithmetic, there is *no* formula that *means* "provability." To say that a formula is provable in a system is to say that there is a finite sequence of steps leading from the axioms to the formula, and the *concept* of a finite sequence is not an arithmetic one. Naturally, we can "encode" every given finite sequence as a single number, just as we can encode each formula of mathematics as a single number, but this does not achieve a reduction of the concept of a finite sequence.

But there are a number of theories—such as set theory—in which one can express the concept of a finite sequence straightforwardly, as Frege showed. We might even argue that the set-theoretical definition of "finite sequence" has become canonical—in the sense that the epsilon-delta definition of continuity in analysis has become canonical. When a formalized notion has become canonical, it no longer makes sense to ask whether the formalized notion "expresses" the intuitive notion. If we consider an "applied" set theory with names for the primitive symbols of the language of mathematics, it is arguable that we know how to write down a formula that not only represents, but actually defines, what Hilbert meant by "provability"; and thus a sentence that paradigmatically expresses the consistency of mathematics. (In first-order number theory, we can encode the finite sequences using such artifices as the "Gödel beta function," but no such encoding can replace the intuitive second-order concept of a finite sequence, which is why we need derivability conditions.) Now Gödel's theorem shows at least that the consistency of set theory cannot be proved (even) in (all of) set theory,[4] and this result decisively refutes the Hilbert program as originally conceived.

The only way out of this for Detlefsen, it seems to me, is to deny that the definition of a proof in set theory is, in fact, "canonical." For the class of proofs is simply the intersection of all classes of formulas that contain the axioms and are closed under the rules of inference. This definition is, of course, impredicative, and there are mathematicians, such as Henri Poincare, who have argued that such definitions cannot replace the intuitions which inspire them.[5]

---

[4] ZF set theory with the addition of the names of the primitive symbols is obviously a conservative extension of ZF.

(b) A more interesting argument put forward by Detlefsen is, in effect, the following: even if the Gödel "translation" of "proof" (and thus "consistency") is accepted to be the only possible translations into arithmetic notation, what Gödel translated was the *realist* notion of proof, not the *instrumentalist* version.

Consider the following concept, which I shall call *Rosser provability,* in honor of its discoverer. Suppose we order all mathematical proofs, say by length, in a theory T extending number theory. Then, relative to this ordering, we can say: a formula is Rosser-provable, if it is provable, and no "shorter" proof refutes it. Let T' be the set of Rosser-provable propositions in the language of T. Of course, if mathematics is consistent, then a formula is provable if and only if it is Rosser-provable: so $T = T'$. On the other hand, even if T is inconsistent, not every formula will be Rosser-provable; and if T is inconsistent, therefore, T' will not be closed under modus ponens.

Now imagine that the "shorter" proofs are also more instrumentally useful than the longer ones. A Rosser-proved proposition, for the instrumentalist, would tend to survive the discovery of a contradiction in T. In other words, we would revise T in order to maintain existing Rosser proofs. Detlefsen speculates that there might be an ordering of the proofs of mathematics such that Rosser proof becomes a concept congenial to Hilbertian instrumentalism, though it has metamathematical properties quite different from the proof concept formalized by Gödel—in particular, Rosser provability violates the derivability conditions.

In fact, we can define "Rosser consistency": T is Rosser-consistent iff one cannot Rosser prove $0 = 1$ in T. But all this condition means is that, whenever the absurdity $0 = 1$ is provable in T, it is refuted earlier on. And thus it is trivial to prove the Rosser consistency of T (i.e., the consistency of T'), assuming that T is consistent. Indeed, it seems plausible that the refutation of $0 = 1$ should rank relatively early in any epistemically significant ordering of proofs, so that the Rosser consistency would become an empty concept.

Now I find myself somewhat sympathetic to the idea that "Rosser proof" is an idealization of some scientific behavior. For example, the mathematical foundations of some quantum field theories are such that nobody knows whether the mathematics is consistent or not. Nevertheless, physicists continue to make calculations, and the idea of "cordoning off" inconsistencies, i.e., refraining from using mathematical instruments where we know they do not work, seems to be applied daily.

---

[5] Cf. Parsons, "Frege's Theory of Number," reprinted in *Mathematics in Philosophy: Selected Essays* (Ithaca: Cornell, 1984).

Still, can "Rosser proofs" replace proofs *in Hilbert's program?* Note that such a replacement would make it, as a practical matter, impossible to prove anything, since one would have to check each and every previous proof to see whether the theorem had been refuted already. Also, even if we could set up an instrumentally useful ordering of proofs, I doubt whether it would correspond to some syntactic ordering (such as length).

But most of all, what has this to do with Hilbert's program? Recall that Hilbert sought to show that "ideal reasoning" would never lead to a false "real" theorem. But a Rosser-proved "real" proposition, in an inconsistent system, might very well be false, as Detlefsen recognizes (140 fn.): the counterexample might appear later in the ordering! Now if we know where the inconsistency lies, we might be able to construct an ordering of proofs where this phenomenon would not exist, or at least be minimized.[6] But in the interesting cases, of course, this does not come into question.

Now Detlefsen is aware of all this, and refrains from claiming that Rosser provability can be used to carry out Hilbert's program (140 fn.). His is the weaker goal of refuting the derivability conditions as a necessary condition on a formalization of mathematical proof. The mere possibility of an ordering relative to which Rosser provability becomes a model for a "strict instrumentalist" concept of proof, he claims, is enough to refute the derivability conditions.

I have already argued that the derivability conditions are not as crucial to the question of whether Gödel refuted Hilbert as Detlefsen and his opponents think. But even if defenders of Hilbert think it vital to refute these conditions, I am mystified why Detlefsen thinks that he can refute these conditions with a proof concept that he himself agrees is not likely to be acceptable to Hilbert! All the defenders of the conditions would have to claim is that they were attempting to capture necessary conditions on any proof concept that could be of use to the Hilbert program (not on any instrumentalistic concept of proof).

Irrespective of Hilbert's program, though, Detlefsen's discussion does suggest that Rosser provability might have its own merit as a model of actual mathematical practice in the physical sciences. Indeed, it might be more interesting if Detlefsen went on to develop his own instrumentalistic philosophy of mathematics than to continue to defend Hilbert's.

MARK STEINER

The Hebrew University/Jerusalem

---

[6] Thanks to Shelah for pointing this out.

philosophy

PHILOSOPHICAL FRAGMENTS
FRIEDRICH SCHLEGEL
Translated by PETER FIRCHOW
Foreword by RODOLPHE GASCHÉ
The work of a philosopher often considered one of history's greatest critics. *$11.95 paper*

THE SAVAGE ANOMALY
THE POWER OF SPINOZA'S METAPHYSICS AND POLITICS
ANTONIO NEGRI
Translated and with a foreword by MICHAEL HARDT
A re-reading of Spinoza's philosophical and political writings. *$18.95 paper*

UNIVERSITY OF
minnesota PRESS
800-388-3863, EXT. 33

THE JOURNAL OF PHILOSOPHY
is pleased to announce the publication of its
TWENTY-FIVE YEAR INDEX, 1964–1988
VOLUME LXI TO VOLUME LXXXV
AUTHORS, SUBJECTS, & REVIEWS
Supplement to its
FIFTY-YEAR INDEX, 1904–1953
TEN-YEAR SUPPLEMENT, 1954–1963

It is available for the rest of 1991 at special prices.

25-year (1964–1988) $25.00 cloth/$10.00 paper
also available
50-year (1904–1953) $12.00 cloth 10-year (1954–1963) $3.50 paper
Shipping $5.00 (first)/$2.50 (each addl.)

Please send a check or money order to

Journal of Philosophy
704 Philosophy Hall
Columbia University
New York, New York 10027

Prices are in effect through December 31, 1991. Orders and payment must be received by then.

COM
[illegible]
KNO
WLE
DGE

Editor
Jeffrey M Perl

Associate Editor
Robert S Nelson

Book Review Editor
Beverly Haviland

Poetry Editor
Belle Randall

Assistant Editor
William Kolbrener

Editorial Board
K A Appiah
Sissela Bok
Cornelius Castoriadis
Stanley Cavell
William M Chace
Robert W Corrigan
Natalie Zemon Davis
Fang Lizhi
Paul Feyerabend
Elizabeth Fox-Genovese, *Chair*
Joseph Frank
Manfred Frank
Clifford Geertz
Eugene Genovese
Ian Hacking
Pierre Hadot
Francis Haskell
Vicki Hearne
Charles Johnson
Stanley N Katz
Hugh Kenner
Sir Frank Kermode
Leszek Kolakowski
Jacques Le Goff
Greil Marcus
Steven Marcus
Adam Michnik
Czeslaw Milosz
Richard Rorty
Abdus Salam
Amartya Sen
Quentin Skinner
G Thomas Tanselle
Charles Taylor
Tatyana Tolstaya
Stephen Toulmin
Gianni Vattimo

Announcement and Call for Papers

The skeptics, neopragmatists, poststructuralists, feminists, Marxists, historicists, Freudians, paleoconservatives, dissidents, and mavericks who have joined to establish **Common Knowledge** hold that fratricidal partisanship is a legacy of a time that is passing. **Common Knowledge** will publish work in cultural studies and intellectual history that redefines divisive terms and figures of the past and present in ways that make expanded sympathies credible. Now that the Berlin Wall is down, it is time to abandon simplistic distinctions– East versus West, left versus right, theory versus praxis, modern versus postmodern It is time to elaborate a basis for practical, common knowledge

Common Knowledge represents a commitment to go beyond established disciplines and, more importantly, beyond the factions at war in the literary and scholarly world. The idea that intellectual life is composed of "issues" and that issues are shaped like battlefields or game boards– with "sides" that one must "take"– is at best obsolete and impractical. **Common Knowledge** seeks to refine a new skepticism– a postmodern, postwar doubt of claims to knowledge and authority– that presents no obstacle to solidarity

Common Knowledge will publish material that focuses on the timely and the particular rather than on the timeless and the theoretical– material that renders common knowledge possible. The editors solicit articles that reassociate intellectual communities divided by the cold war, that establish links among disparate intellectual schools, that deepen the conversation within each school, that break down the distinction between scholarly and narrative form, that reopen on higher ground debates about incommensurability, ambivalence, indeterminacy, untranslatability, and the will-to-power

Manuscripts, proposals for articles, and books proposed for review should be sent to **Common Knowledge**, School of Arts and Humanities, The University of Texas at Dallas, Box 830688-JO 31, Richardson, TX 75083-0688. For subscription information, contact the Journals Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513

PROMETHEUS
BOOKS

Animal Experimentation The Moral Issues
edited by Robert M. Baird and Stuart E. Rosenbaum

This balanced collection of 16 articles and essays—the newest addition to Prometheus's "Contemporary Issues" series—captures the clash of opposing positions in the national debate over animal rights and animal welfare. The intensity of this conflict is sure to increase as demands for more research make the need for informed discussion all the more urgent. Contains selections by Carl Cohen, Alan Freeman, J. A. Gray, Peter Harrison, Edwin Hettinger, Betty Mensch, Tom Regan, Bernard E. Rollin, Richard Ryder, Richard H. Schwarz, Peter Singer, William Timberlake, Mary Anne Warren, Robert B. White, Robert Wright, and Steven Zak.

182 pages • Paper $14.95

Animal Consciousness
Daisie Radner and Michael Radner

Questions about animal consciousness are often dismissed as unanswerable. Here the authors address many complex issues from varying perspectives: the history of philosophy, the philosophy of science, the history of science, and the philosophy of mind. Drawing upon the literature of philosophy, psychology, biology, and ethology, they challenge the doctrines that have led to an impasse in the human understanding of animals. The Radners argue that the usual reasons given to undermine the validity of animal consciousness as a subject of scientific inquiry—that it is ill-defined and superfluous—are themselves ill-founded. They conclude by exploring the ethical implications of giving serious consideration to animal consciousness.

"*Well-constructed and adroitly argued.*" —*Nature*

253 pages • Cloth $39.95

Animal Rights and Human Morality
Bernard Rollin

A lucid analysis of animal rights that includes numerous examples of how research on animals could be dramatically improved without detriment to human welfare.

"*A relaxed tone and personal anecdotes make the book fascinating for general readers; the depth of discussion makes it valuable for philosophers and animal welfare workers.*" —*Library Journal*

182 pages • Cloth $23.95 • Paper $15.95

Order Toll Free (800) 421-0351 (24 hours)
or write to
59 John Glenn Drive • Amherst, NY 14228-2197
Add $3.00 p&h for first book, $1.50 for each additional book ($9.00 maximum)
NY State residents add sales tax.
#40114

NEW TITLES

Apocalypse and/or Metamorphosis

NORMAN O. BROWN

"Few of the masters of modern consciousness have continued to develop as Brown has done. . . . The contents [of this book] are dazzlingly original. . . . Now, here, in 1990, he looks beyond the 'post-modern' condition." —Hayden White, author of *Metahistory*

$22 50 cloth

The Grain of the Voice

Interviews 1962-1980

ROLAND BARTHES

Translated by Linda Coverdale

Reissued in paperback

"Perhaps the best introduction to date on Barthes and his work."

—*The Bloomsbury Review*

$12 95 paper

Observations on the Feeling of the Beautiful and Sublime

IMMANUEL KANT

Translated and Edited by John T. Goldthwait

New in paperback

"This early aesthetic work of Kant . . is clearly and simply written and shows a deep understanding of human nature. "—*Philosophical Books*

$10 95 paper

Heritage of Our Times

ERNST BLOCH

Translated, with an Introduction by Neville Plaice and Stephen Plaice

Recalling work by Walter Benjamin and the Frankfurt School, Bloch's study of everyday life and politics during the Weimar Republic and Nazi Germany is a brilliant historical analysis of the cultural conditions leading to fascism.

Weimar and Now
German Cultural Criticism

$40 00 cloth

The Responsibility of Forms

Critical Essays on Music, Art, and Representation

ROLAND BARTHES

Translated by Richard Howard

Reissued in paperback

"Like Freud before him, the later Barthes seeks to install himself at the very juncture of sign and body, semiotic and somatic, meaning and materiality "

—*Times Literary Supplement*

$12 95 paper

The Boundaries of Humanity

Humans, Animals, Machines

Edited by
JAMES J. SHEEHAN and MORTON SOSNA

New in cloth & paperback

"Rethinking our relationship to animals is very relevant, I believe, to thinking clearly about our current relationships to current (and future) machines."—Keith Gunderson, University of Minnesota

$35.00 cloth, $12.95 paper

Whose Keeper?

Social Science and Moral Obligation

ALAN WOLFE

New in paperback

"A book that offers genuinely fresh insights into the ever more intractable dilemmas of modernity."—*Commonweal*

Winner of the C. Wright Mills Award

$12.95 paper

Women and Evil

NEL NODDINGS

New in paperback

"The power of Noddings' book . . . lies in its challenge to traditional ethical systems. . . There is something sensible, immediate, and profoundly thought-provoking about her claims." —*Theological Studies*

$12.95 paper

Nishida Kitaro

NISHITANI KEIJI

Translated by Yamamoto Seisaku and James W. Heisig
With an introduction by D. S. Clarke

Nishitani wrote these essays about his teacher from 1936 to 1968. A series of meditations by one master on another, this book provides a remarkable, living portrait of Nishida Kitaro.

Nanzan Studies in Religion & Culture *$25.00 cloth*

Democracy and Moral Development

A Politics of Virtue

DAVID L. NORTON

"The author is able to bring philosophical ideas to bear upon actual problems of politics in a most refreshing way." —Richard Taylor, author of *Good and Evil*

$29.95 cloth

The Direction of Time

HANS REICHENBACH

Edited by Maria Reichenbach
With a new introduction by Hilary Putnam
New in paperback

This far-sighted work illuminates the metaphysical questions with which researchers in cosmology and particle physics must contend.

$14.95 paper

At bookstores or order toll-free 1-800-822-6657 Visa/MasterCard only

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY LOS ANGELES NEW YORK OXFORD

Cambridge University Press

New in paperback…

The Economy of the Earth
Philosophy, Law and the Environment
Mark Sagoff

"His book serves as an outstanding example of how applied philosophy should be done."

—*Times Higher Education Supplement*

"Any person pursuing a graduate degree in natural resource economics with interests in public policy analysis needs to read and reflect upon Sagoff's book. Any practicing natural resource economist should do the same. This is an important book."

—*The American Journal of Agricultural Economics*

Cambridge Studies in Philosophy and Public Policy
39566-6 Paper $12.95

New in paperback…

The Philosophy of Quantum Mechanics
An Interactive Interpretation
Richard A. Healey

"In a very careful, subtle, and deeply original piece of work, Healey has achieved something no one else has been able to do in the sixty-year history of the discussion of quantum mechanics."

—*Howard Stein, University of Chicago*

40874-1 Paper $12.95

New in paperback…

Forgiveness and Mercy
Jeffrie Murphy and Jean Hampton

"Their analyses of these crucial emotions are clear, elegant and sufficiently controversial to initiate an exciting and much needed discussion of the proper role and the nature of the personal passions in social practice and theories."

—*Robert C. Solomon, University of Texas, Austin*

"…an impressive book. Anyone who is interested in the topic ought to read it."

—*Ethics*

Cambridge Studies in Philosophy and Law
39567-4 Paper $10.95

Autonomy and Self-Respect
Thomas E. Hill, Jr.

"The book would be an invaluable accompaniment to Kant's own texts in a Kant's ethics course and it would make a much more interesting main text than one can usually get in a practical ethics course."

—*Christine Korsgaard, University of Chicago*

39464-3 Hardcover $39.50

Contractarianism and Rational Choice
Essays on David Gauthier's Morals by Agreement
Peter Vallentyne, Editor

A number of prominent moral and political philosophers offer an assessment of Gauthier's theory—the most complete and suggestive contractarian theory of morality since the work of Rawls—and he responds in turn to their "critique."

39134-2 Hardcover $59.50
39815-0 Paper $19.95

Philosophy in Christianity
Godfrey Vesey, Editor

An examination of the central doctrines of Christianity and the ways in which Christians have drawn on philosophical ideas in the formulation of these doctrines.

Royal Institute of Philosophy Lecture Series 25
37678-9 Paper $16.95

Philosophy and Politics
G. M. K. Hunt, Editor

A volume of original works explores the relationship between the nature of man and the structure of society as part of a recurrent theme connecting philosophy and politics.

Royal Institute of Philosophy Supplements 26
39697-6 Paper $16.95

Available in bookstores or write:

CAMBRIDGE UNIVERSITY PRESS
40 West 20th Street, New York, NY 10011-4211. Call toll-free 800-872-7423.
MasterCard/VISA accepted. Prices subject to change.

CENTRAL LIBRARY

19 SEP 1991

THE JOURNAL OF PHILOSOPHY

philosophy

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

VOLUME LXXXVIII, NUMBER 7
JULY 1991

page
337 *Reductive and Nonreductive Simile Theories of Metaphor* Lynne Tirrell
359 *Reference and Pronominal Descriptions* George M. Wilson
388 NEW BOOKS

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1991 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NO. 7, JULY 1991

REDUCTIVE AND NONREDUCTIVE SIMILE THEORIES OF METAPHOR*

METAPHOR is often defined as implicit or indirect comparison. Those who say this generally are committed to the possibility of reducing metaphor to simile (direct comparison) on the grounds that both are comparisons and with the belief that we can make better sense of simile than metaphor. According to this line of thought, our account of metaphor should be parasitic on our account of simile, since metaphor is just subtle simile. The argument, made explicit, has two main steps: reduce metaphor to simile, then conflate simile and literal comparison. 'Juliet is the sun' is nothing other than a shorter, neater, more powerful version of 'Juliet is like the sun', and to understand the former we must first understand the latter. The simile theory attempts to save literal semantics for metaphor by reducing metaphor to simile, and only succeeds if a conflation of simile and literal comparison is allowed.

Most simile theorists hold that on the surface 'Juliet is the sun' is obviously false, since Juliet is a young woman and one physical object cannot be simultaneously both a young woman and the sun. On the other hand, they point to the fact that it looks as if 'Juliet is like the sun' is trivially true. One advantage of simile over metaphor is that similes are usually if not always true.[1] (Donald Davidson, for example, claims that similes are always true because "everything is like everything else.") If we take metaphor to be a special case of simile, then we can see why a sentence that on most semantic analy-

---

* I would like to thank Robert B. Brandom, Joseph L. Camp, Jr., William G. Lycan, Stanley Munsat, Michael Resnik, Jay F. Rosenberg, Geoffrey Sayre McCord, and Josef Stern for helpful comments on an earlier draft of this paper.

[1] Robert Fogelin denies this claim, on the grounds that their simply sharing a property is not sufficient for saying that two objects are similar. See his *Figuratively Speaking* (New Haven: Yale, 1988), p. 59 ff.

0022-362X/91/8807/337–58

ses is false may nevertheless "ring true."[2] And we may do so without recourse to special metaphorical or figurative meanings.[3]

This view—the reductive simile theory of metaphor—underlies many current theories of metaphor and claims for itself several virtues. First, it resists appeal to multiple kinds of meaning, by relying solely on literal meaning (plus a few rules) to determine semantic content. Second, it holds out a greater promise for a systematic treatment of figurative language if figurative language can be explained in terms of the literal, since so many philosophers of language think literal talk can be explained compositionally. Third, the reductive simile theory makes sense of our attempts to understand specific metaphors. Upon encountering a new, somewhat obscure metaphor, say 'Juliet is the sun', we do not ask whether Juliet is or is not the sun but rather ask how Juliet is *like* the sun. Thus, the simile theorist claims, we intuitively reduce metaphor to simile every time we try to understand a particular metaphor. To find the meaning of a metaphor one finds the meaning of its corresponding simile. To find the meaning of the simile, one finds the meaning of the corresponding literal comparison.

Not all simile theories of metaphor are so explicitly reductive. Recently, Robert Fogelin has presented a defense of a nonreductive simile theory in his *Figuratively Speaking*. Fogelin's attempt to save the simile theory avoids the explicit reductionism that has characterized the theory in the past, but leaves us with the question of what use there is in assimilating metaphor to simile after all. After showing what is wrong with the standard simile theory of metaphor, which reduces metaphor to simile and simile to literal comparison, I shall address Fogelin's less thoroughly reductive position.

I REDUCTIVE SIMILE THEORY

Although the details of the reductive simile theory may vary widely, certain claims define the theory. Claim (0) (below) is endorsed by most simile theories, but it is not strictly necessary.[4] The conjunction of (1), (2), and (3) is entailed by any reductive simile theory in virtue of being such a theory.

---

[2] Upon analysis, we shall see that rather than metaphor be a special case of simile, simile is a special case of metaphor. In *Languages of Art*, Nelson Goodman briefly remarks: "Instead of metaphor reducing to simile, simile reduces to metaphor; or rather, the difference between simile and metaphor is negligible"—(Indianapolis: Hackett, 1976), p. 77–8. I call simile a "special case" of metaphor rather than claim a reduction, because I do not want to go so far as Goodman's latter clause.

[3] In their adjectival forms ('metaphorical', 'figurative'), these terms are usually treated as equivalent, even though the nominal forms are not so treated. For example, 'metaphor' admits of both a generic interpretation, according to which it is equivalent to 'figure', and a specific interpretation, which denotes a particular sort of figure and no others. Since the adjectival forms are usually taken to denote the genus, I shall use them accordingly as interchangeable.

(0) Simile and metaphor play the same linguistic role: they make (or prompt us to make) comparisons.
(1) Any metaphor can be paraphrased as an identity or predication.
(2) For any metaphor of the form '*A* is *B*' or '*A* is a *K*', there is a function from that metaphor to a simile.
(3) The similes that result from (2) can be explained in purely literal terms.

In one form or another, this view is shared by such diverse writers as Cicero, Davidson, Paul Henle, Hegel, S. J. Brown, George Miller, Max Black, and Scott Buchanan, plus the myriad writers of textbooks on English composition and on rhetoric.[5]

The simile theory is interesting as a case of reduction in the philosophy of language, but it is also interesting because it underlies so many diverse theories of metaphor. Consider Black and Davidson, for example. Neither is usually found on a list of simile theorists. Black attempts to explain how metaphors get their special figurative meaning (really a sort of pragmatic meaning), while Davidson tries to show that metaphor has no special figurative meaning whatsoever.[6] Davidson even argues against several versions of the simile theory, and only emerges as a simile theorist at all once attention is paid to his view of the relation of semantics and pragmatics to metaphor. For Davidson, the literal semantic meaning of the simile is the basic pragmatic meaning of the metaphor. This is not so far from Black, after all, since Black claims that simile provides an adequate paraphrase of metaphor semantically but would deny that metaphor and simile play the same pragmatic role. Some writers, like Cicero, Buchanan, and Hegel, explicitly endorse the simile theory, without making explicit each step of the reduction. Many current writers on metaphor, however, explicitly deny the simple simile theory while implicitly holding a more sophisticated version, usually one that makes important use of the troublesome distinction between semantics and pragmatics.

---

[4] Designating the claim as '0' is meant to set it apart from (1)–(3), since it is usually (but not always) treated as an assumption.

[5] See Cicero, *De Oratore*, H. Rackham, trans. (Cambridge. Harvard, 1942), Davidson, "What Metaphors Mean," in Sheldon Sacks, ed., *On Metaphor* (Chicago: University Press, 1978), pp. 29–45; Henle, "Metaphor," in *Language, Thought, and Culture* (Ann Arbor: Michigan UP, 1965), pp. 173–95; Hegel, "The Metaphor, the Image, and the Comparison," in *The Philosophy of Art*, William M. Bryant, trans. (New York: Appleton, 1879), pp. 40–5; Brown, *The World of Imagery* (London: Kegan Paul, Trench, Trubner, 1927); Miller, "Images and Models, Similes and Metaphors," in A. Ortony, ed., *Metaphor and Thought* (New York: Cambridge, 1979), pp. 202–50; Black, "More About Metaphor," in *Metaphor and Thought;* Buchanan, *Poetry and Mathematics* (New York: John Day, 1929). Unlike the other writers noted, Black's endorsement of (0) is limited.

[6] Pragmatic meaning is roughly the interpretation generated after attending to both the strictly semantic (context-independent) meaning of the expression interpreted literally and aspects of the context of use which support or require a reinterpretation of the expression. Figurative meaning is one sort of pragmatic meaning.

My argument against the simile theory is quite simple. Simile theorists emphasize claim (2), usually assuming (1) and (3), with (0) in the background as the most basic assumption. My strategy is to show that neither (1) nor (3) is true, and that, without (1) and (3), (2) is of little help in developing an account of metaphor. With (1), (2), and (3) undermined, there will be no reason to grant (0) as an assumption. (There are independent reasons for rejecting (0), which will become apparent later.) Let us begin by noting one obvious problem with claim (2).

I.1 Claim (2).

(2) For any metaphor of the form '*A* is *B*' or '*A* is a *K*', there is a function from that metaphor to a simile.

The simile theorist is quick to point out that generating a simile that corresponds to a particular metaphor is often straightforward.[7] From 'Juliet is the sun' we make 'Juliet is like the sun', from 'Richard is a lion' we make 'Richard is like a lion', from 'Old fools are babes again' we make 'Old fools are like babes again', and so on. For any metaphor of the form '*A* is *B*' or '*A* is a *K*' we first try the basic algorithm: let a 'like' follow the copula. This does seem straightforward. The simile gives us a sentence to analyze which is not obviously false and provides a structure that informs us that the meaning is comparative.[8]

---

[7] This claim is limited to those metaphors of the required form, and so is only of general use in a theory of metaphorical or figurative discourse if claim (1) is true.

[8] If we distinguish between the semantic role of an expression and its pragmatic role, we can see that adherents of claim (2) usually maintain either that

2.1.a. For any metaphor, there is a function from the semantic role of its corresponding simile to its own semantic role. (This function may or may not be identity.)

and

2.1.b. There is a function from the semantic role of the metaphor to the pragmatic role of the metaphor.

The simile does seem to be comparative; what is not clear is that the original metaphor is. Adding a semantic marker is not trivial. No one would argue that, since we can take any claim of the form '*A* is a *K*', add a 'not' after the copula, and still have a syntactically correct and semantically meaningful sentence, the deep grammar of all predications is negation. What is obvious in the case of adding 'not' is less obvious (but still true) in the case of adding 'like'. When interpreting literally, there is a big difference between saying '*A* is a *B*' and saying that '*A* is like a *B*'. The former entails *A*'s inclusion in the set of *B* things, while the latter does not. The 'like' weakens the semantic commitment of the original sentence.[9] Since the semantic commitments of the two claims differ, the addition of the 'like' is not trivial. Why should we think that it is trivial when interpreting figuratively? I shall come back to this question.

I.2 Claim (3).

(3) The similes that result from (2) can be explained in purely literal terms.

To say what 'Juliet is the sun' means, according to the reductive simile theory, we must explain the respects in which Juliet is like the sun. The relevant meaning of the simile is an ***interpreted*** meaning; we must say what inferences the expression sanctions when read figuratively. Juliet is warm, she is bright and dazzling, she is the center of Romeo's world, his day begins with her, and so on. (Does she, like the sun, burn alive those men who draw too near?) Still, we are no nearer the literal, with its promise of ready-made semantics.

'Juliet is warm' is metaphorical in a way that 'The sun is warm' is not. (The vagueness of 'warm' need not concern us.) The same holds for 'Juliet is bright'. (It is silly to say "Get out your sunglasses; Juliet is bright today!") 'Romeo's world revolves around her' does not mean that the patch of earth upon which Romeo stands revolves around Juliet as the earth revolves around the sun; rather, it means

---

or that

2.2 There is a function from the semantic role of the simile to the pragmatic role of the metaphor.

Both 2.1 and 2.2 take us from the semantic role of the simile to the pragmatic role of the metaphor. 2.1 does this by way of the semantic role of the metaphor; 2.2 does not.

[9] The 'like' weakens the claim to which it is applied because the inferences that one may draw from *A*'s being a *B* are no longer justified by the claim that *A* is like a *B*. The weakening function of 'like' is discussed further in section II.3.

that Romeo places central importance on Juliet's happiness and well-being and has made her his raison d'être. *Explication of the simile leads us to more metaphors.* Not only that—the explication of any simile leads us to metaphors, and has to.

To see why explicating simile leads to metaphors, let us compare simile (either live or dead) to literal comparison. Recall that much of the appeal of the simile theory lies in its purported ability to convince us that there is no significant gap between the figurative and the literal, at least concerning metaphor. The deep structure of metaphor (the paradigm of figurative discourse) is simile, and simile can be understood wholly literally. Let us begin by addressing the naive-realist version of the literality thesis. Then, I shall show that some similes cannot be understood wholly literally (thereby undermining the universal claim of the simile theorist), and finally we shall see why no similes may be understood wholly literally.

First we must note that the simile theory is often accompanied by a naive-realist version of the literality thesis. According to the naive realist, in order to find the meaning of 'Juliet is (like) the sun', one finds Juliet, finds the sun, and compares them. The similes mentioned so far are all easily taken as having ordinary physical objects as referents of the *A* term and the *B* term. Not all similes are so cooperative. Although it is possible to consider the warmth and brilliance of the sun in trying to understand what it means to say Juliet is warm and brilliant (possible, although not very fruitful), what are we to examine when we try to understand T. S. Eliot's "He laughed like an irresponsible foetus"? Assuming, for the moment, that we know to whom the 'he' refers, how can we find an irresponsible foetus whose laughter we can study?[10] We cannot, since there is no such thing. Eliot augments the simile and gives clues for its interpretation by way of metaphor and further simile, saying

His laughter was submarine and profound
Like the old man of the sea's
Hidden under coral islands
Where worried bodies of drowned men drift
    down in the green silence,
Dropping from fingers of surf.

(from "Mr. Apollinax")

---

[10] This may seem to be a contrived objection, but see George A. Miller, "Images and Models, Similes and Metaphors," for a particularly explicit version of this view (in A. Ortony, ed., *Metaphor and Thought*, pp. 202–50. As John Searle points out (in "Metaphor"), most comparison theories are sufficiently Aristotelian to make such existential presupposition; such presupposition precludes an adequate account of similes containing nondenoting expressions.

Our task here is not to determine the exact meaning of the simile but to note that we cannot expect that for any simile we need merely to find the objects to which the terms refer and then compare them. A naive-realist version of the simile theory cannot handle all the similes we can understand, much less all the metaphors. In particular, it cannot handle similes containing nondenoting expressions.

Not all similes can be understood wholly literally. There is, as we have just seen, the problem of similes containing expressions that fail to refer (in the ordinary sense of 'refer'). A Dutch expression used to praise especially fine food may be translated as "It's like angels pissing on your tongue."[11] Although angel-lore omits reference to pissing (and one would expect that angels do not) and although there are not really any angels to do anything even like pissing, the simile does not say "This tastes like nothing doing what it cannot do." Waxing compositional does not help. Failure to refer is not the mark of the figurative, any more than falseness is. 'Angel' has both a literal and a figurative or metaphorical interpretation. If I tell my three-year-old niece that her dead dog is an angel now, I have used the term literally even though it fails to refer. If I tell her to be an angel and not interrupt, I have used the term figuratively and told her to be well-behaved, not to drop dead.

The fact is that no simile can be understood wholly literally. A sentence of the form '*A* is like *B*' is not ipso facto a simile. In 'A bicycle is like a tricycle except a bicycle has two wheels instead of three', 'a bicycle is like a tricycle' is not a simile. It is a straightforward literal comparison. I can justify my assertion of 'The engine of my boat is like the engine of my car' with a long tedious list of properties both have—in the same literal sense. (Both are physical objects, made of metal, plastic, and rubber, both power vehicles, both are of internal-combustion design, both run on gas, and so on.) Although of the correct syntactic form, such a comparison is not a simile. Neither the inferences it licenses nor the justification it requires involves figurative discourse in any way.[12]

---

[11] Thanks to Arthur Ripstein for introducing me to this expression.

[12] I am operating with an inferential role theory of meaning. I distinguish between three basic sorts of commitment associated with linguistic practice: identificatory, assertional, and expressive We make commitments in referring, asserting, and invoking metaphors because we may be called upon to identify the referent of our terms, justify the assertional content of our claims, or make appropriate use of extensions to explicate metaphorically interpreted discourse. For a discussion of assertional commitment, see Robert B Brandom, "Asserting," *Nous*, IV (November 1983): 637–50. For a discussion of asserting metaphors, see Merrie Bergmann, "Metaphorical Assertion," *The Philosophical Review*, XCI (April 1982): 229–46. For a fuller discussion of extended metaphor and expressive commitment, see my "Extending The Structure of Metaphor," *Nous*, XXIII (March 1989). 17–34

The widespread reliance on some sort of semantic marker (e.g., 'like', or 'as') to mark the distinction between simile and metaphor is possible because most theorists define simile as a comparison with a semantic marker, then define metaphor as simile with the marker left out, and then conflate the definitions. Explicitly to define simile as metaphor with a semantic marker and metaphor as simile without the marker would be circular, of course, but by failing to so much as wave their hands in the direction of a distinction between simile and literal comparison these theorists implicitly rely on such a circle. Once we recognize, however, that in moving from metaphor to simile we have not moved from trope to nontrope, from figurative to literal discourse, it is clear that the simile theorists have not completed their reductive program but, at best, relocated it. They still owe us a strict reduction from simile to literal comparison, but that is a reduction from simile to a class they do not and arguably cannot define.

Often those who try to define simile syntactically claim that for simile, unlike literal comparison, the *A* term and the *B* term must represent different classes. The simile is taken to be a comparison between two unlike objects, two objects in "essentially different categories,"[13] or two "distinctly different things."[14] Even assuming that we know whether two given terms refer to items in different classes, and most nonidentical terms do, this requirement will not distinguish between similes and literal comparisons. Some sentences (sentence types) of the correct form, with the cross-class requirement fulfilled, can be used either as a literal comparison or as a simile. Consider 'George is like the dog next door'. The sentence is a simile if it licenses inferences or requires justifications such as "Both bark but neither bites," but it is not a simile if the relevant response is "They both have brown hair, both weigh 220 pounds, both have big teeth and bad breath," and so on.[15]

The presence of a semantic marker may distinguish simile from metaphor if we are sure the expression in question is a trope, but it cannot distinguish simile from literal comparison. The cross-class requirement cannot distinguish simile from literal comparison, since even if we grant that the terms denote objects belonging to different classes ('human', 'dog') we can give the same sentence both a literal reading and a figurative reading. (And notice here that 'figurative' is just another way of saying 'metaphorical'.) Thus, the distinction between simile and literal comparison lies in the sort of interpretation an audience assigns (or ought to assign) to the expression.

---

[13] Raymond W. Barry and A. J. Wright, *Literary Terms: Definitions, Explanations and Examples* (San Francisco: Chandler, 1966).

[14] M. H. Abrams, "Figurative Language," *A Glossary of Literary Terms* (New York: Holt, Rhinehart & Winston, 1957), pp. 60–3.

[15] Actually, even this is overstated, since we may interpret 'barks' and 'bites' either literally or figuratively.

I.3 Claim (1).

(1) All metaphors can be reduced to or paraphrased as predicates or identities.

We have already discussed the process of generating a corresponding simile for a metaphor of the simple identity form. It is usually easy to apply the basic algorithm to sortal predication metaphors, such as 'Fran is a fox'. These are the only cases the reductive simile theorists discuss, and then not under these descriptions. Usually they take one or the other form as the paradigm for all metaphor.[16]

We have been talking as if all metaphors work the same way, but this has not been shown and is not a trivial assumption. *If* the various forms of metaphor cannot all be reduced to either predicates or identities (or to at least one of these two categories), then the claim (2) reduction briefly discussed earlier will be idle. Although it may be true that any assertion can be put into the form '*A* is *B*', it is not true that any metaphor can be. First, not every metaphor is an assertion. Second, and equally important, even if we make every metaphor into an assertion, we cannot always get the right terms into the *A* and *B* positions. Once these two points are established, we shall see that the simple transformation of predications and identities from metaphor to simile cannot carry the theoretical weight it has been assigned.

To see that not every metaphor is an assertion, one need simply look at logically complex metaphors, as in the following lines from Yeat's *Sailing to Byzantium:*

> An aged man is but a paltry thing,
> A tattered coat upon a stick, unless
> Soul clap its hands and sing, and louder sing
> For every tatter in its mortal dress,
> Nor is there singing school but studying
> Monuments of its own magnificence;
> To the holy city of Byzantium.

In fact, this stanza is full of metaphors that are not assertions. (This is not quite the same as saying that they are not asserted.) Taking account of the logical form of the first few lines above, the following assertion can be attributed to Yeats:

---

[16] Even Fogelin does this; see *op. cit.*, p. 97.

> If the soul does not clap its hands and sing and louder sing for every tatter in its mortal dress, then an aged man is but a paltry thing, a tattered coat upon a stick.

Although we have an assertion, none of the metaphors is asserted since they are embedded in the conditional.

Next, to show that, even if one could make every metaphor into an assertion, one cannot always get the right terms into the *A* and *B* positions, we need to look at a variety of types of metaphor. There are at least six kinds of expressions which may be interpreted metaphorically, and it is not obvious that they can be collapsed into one. First, there are the ***simple identities*** of the form '*A* is *B*', where '*B*' is a name or definite description. This is the most frequently discussed metaphor form, and 'Juliet is the sun' is an instance of it. (Metaphors that make an indefinite description from a name, such as 'He is a Casanova', do not fit this category.) Next, there are ***pure predications,*** of the form '*A* is ϕ', such as 'Juliet is warm' and 'Juliet is brilliant'. Third, there are ***sortal predications,*** of the form '*A* is a *K*', which seem to combine the first two categories. Do we analyze 'Fran is a fox' as saying that there is a fox that is Fran or as saying that Fran has certain traits that generally characterize foxes? These first three types of metaphor do not pose obvious problems for the syntactic maneuvers required by the simile theorist. Similar to predications and identities are ***substitution metaphors*** achieved by substituting a term that does not literally apply for one that does, as in e.e. cummings's

> pity this busy monster, manunkind
> not

This sort of metaphor seems quite close to speaker reference, since it is clear that cummings uses 'monster' to refer to humanity.[17]

Most troublesome for the simile theory are metaphors that operate as functions. There are the ***noun-function metaphors,*** of the form 'The *B* of *A*', which apply to an object *A* a function *B* that does not usually have a value for that object. Examples include Blake's "The countless gold of a merry heart, The rubies and pearls of a loving eye," and sentences like 'Argentina is the England of Latin America'. It is not crucial that the article be definite: 'a cloak of silence' is also a noun-function metaphor. Finally, there are the *verb-function metaphors.* (These are perhaps the most common metaphors.) These are of the form 'The *A Xes*', where *X* is something *Bs* typically do. It is at least prima facie tenable to maintain that something like '*A* is *B*' is presupposed by such a metaphor. Examples of this form include Blake's "When the green woods laugh with the voice of joy," Eliot's "The river sweats/oil and tar," and Dickinson's "Frequently the hills are pink,/Frequently are brown./Frequently the hills undress/Behind my native town."

---

[17] Searle's account of metaphor is well-suited to substitution metaphors and any others that mimic speaker reference.

The corresponding similes for metaphors that are neither predications nor identities often involve quantificational terms. The basic algorithm is insufficient for generating a simile from Blake's "The rubies and pearls of a loving eye," but with some extra work we can get 'whatever a loving eye has that is like rubies and pearls'. The 'whatever' is quantificational and stays quantificational, because there is no object (however fictional, abstract, or linguistically constructed) to which we can attach a name and so generate a substitution instance of the quantified formula.[18] By making this transformation, we gain generality at the cost of accuracy and informativeness.

The same problem faces the verb-functional cases. Take, for example, one sportscaster's hyperbolic verb-function metaphor: 'When Sandy Koufax was young he could throw a strawberry through a locomotive'.[19] There are two ways to generate a simile from this sentence, and both involve the introduction of quantificational terms. The simile theorist might say:

(a) When Sandy Koufax was young he was like someone who could throw a strawberry through a locomotive.

or

(b) When Sandy Koufax was young he could do something like throw a strawberry through a locomotive.

(a) quantifies over people; (b) quantifies over actions. If we want our quantified simile to be close to the English metaphor, we shall pick (b). That is, unless we decide that every metaphor of the form 'The *A Xes*' is to be understood as presupposing a sentence of the form '*A* is *B*' and that the sentence presupposed is the key to the meaning of the verb-functional metaphor. In that case, we would choose (a) and attempt to reduce this category to the category of simple identities. This second alternative explicitly follows the simile theory's reductionist strategy. The sportscaster, however, was not saying that Sandy Koufax was ***like*** someone who could throw a strawberry through a locomotive and that if we want to know what Koufax was like we should find such a person. We have already dismissed the naive-realist version of the theory. Treating the simile quantificationally leaves us with a permanent quantifier.

---

[18] Since we have already eliminated the naive-realist version of the simile theory, it is not important that our substituends be physical objects. On any account of quantification this quantifier remains.

[19] Thanks to Joe Camp for this example.

This brief look at a variety of types of metaphor has shown that generating an '*A* is *B*' sentence with the appropriate terms in the *A* and *B* positions is not always a trivial task. Most troublesome are the cases that leave us with ineliminable quantifiers in the "corresponding" simile.

A story about how we could determine the meaning of a metaphor can be told using the move to a quantificational simile, but a better story can be told without it. To return to the Koufax example, if the metaphor means that Sandy Koufax is like someone who can throw a strawberry through a locomotive, it does not really matter whether there is or is not such a someone. ***What is important is what such a someone would be like, what traits he or she would have.*** The problem with the simile approach to such cases is that it introduces unnecessary complications, sending us in the wrong direction, looking for properties shared by (in this case) Koufax and some nonexistent someone who really has the properties attributed to Koufax in the metaphor. We need to use the dummy subject to generate a simile, but to understand the metaphor we can and do skip the quantificational stage and ask, "What would Sandy Koufax have to be like in order to have such a capacity?" And we answer, he would have to have a frightfully fast arm and a keen eye. Such an answer provides a basis for understanding the metaphor.

The simile theorist will also have to use quantification in generating similes from metaphors involving pure predication. Given that the simile theorist is determined to make every metaphor into a comparison, 'Juliet is warm' will have to become 'Juliet is like a warm thing' or 'Juliet is like something warm'.[20] Unless one is from the San Fernando Valley, 'Juliet is like warm' will not do. The quantificational terms get the sentence into the right form, but they not only fail to add anything that can help us to understand the metaphor, they also weaken the sentence. Now we no longer attribute warmth to Juliet but rather something like warmth! Simile waters down the assertional commitment of a purely predicative metaphor.

---

[20] One might be tempted to deny that 'Juliet is warm' is a metaphor at all, since one of the dictionary's entries for 'warm' is 'cordial', which may be applied perfectly literally to Juliet. It is crucial to note that (a) not every dictionary entry represents a literal interpretation of the term being characterized, and (b) although 'Juliet is cordial' may most readily support literal interpretation, its purported synonymy with 'Juliet is warm' does not show that the latter should or must be interpreted literally in the context in question.

The reductive simile theory is untenable. It fails primarily because the reduction of simile to literal comparison fails. It also fails because the transformation of metaphor to simile does not always preserve meaning. In addition, we have seen that there are metaphors that cannot be paraphrased as an identity or predication (contra claim (1)), and this shows the limited value of the simile theorist's reliance on simply adding a 'like' after the copula. Further, the differences in interpretations of supposedly corresponding metaphors and similes which have been noted in this section give us reason to doubt that simile and metaphor serve the same linguistic purpose (contra claim (0)). Since both the reductive and the nonreductive simile theorist hold that the shared purpose of these tropes is to make or to prompt us to make comparisons, let us examine the nonreductive simile theory before explicitly addressing claim (0). After reviewing the nonreductive simile theory, we should be in a better position to see whether metaphor and simile play the same linguistic role.

II. NONREDUCTIVE SIMILE

A less reductive simile theory of metaphor has been presented recently by Fogelin in *Figuratively Speaking*. Fogelin's characterization of the simile theory is slightly different from mine, primarily because he sets out a view that accepts claims (0), (1), and (2) but rejects (3). Fogelin does not make the mistake of assimilating simile and literal comparison, and that is why I call his view *nonreductive*. Like other simile theorists, he does attempt to reduce metaphor to simile, and so the criticisms of (1) and (2) presented above do apply to his view. Most important to Fogelin is the claim I have identified as claim (0), which contends that simile and metaphor do the same work: they make or prompt us to make comparisons.

On Fogelin's account, the simile theory is characterized by two theses, which he accepts:

> I. The *literal* meaning of a metaphor of the form 'A is a ϕ' is the same as the *literal* meaning of the counterpart simile of the form '*A* is like a ϕ'.
>
> II. The *figurative* meaning of a metaphor of the form 'A is a ϕ' is the same as the *figurative* meaning of the counterpart simile of the form 'A is like a ϕ' (*op. cit.*, p. 29).

Thus, according to Fogelin's view, given a metaphor of the form '*A* is a ϕ', it literally and figuratively means that '*A* is like a ϕ'. Although Fogelin thinks that the second thesis depends upon the first, in what follows I shall treat them separately. The first thesis is easily undermined, as we shall see, and it would be cheap and uninformative simply to dismiss the second thesis because the first fails. The problem with the first thesis does not directly undermine the second thesis, and so the second thesis should be addressed independently. As stated in *Figuratively Speaking,* Fogelin's position does not support the second thesis, but we shall see that attending to that position presents suggestions for the development of a more systematic treatment of figurative interpretation. Let us look first at how Fogelin distinguishes simile from literal comparison, for it shows how he characterizes the figurative, and then turn to what he says about simile and metaphor.

II.1 Simile and literal comparison. On Fogelin's view, the first main difference between simile and literal comparison is that simile arises from a special sort of interpretation on the part of both speaker and audience, and can be assimilated to indirect speech acts. In particular, he appropriates a Gricean speech-act framework. Although we shall see that this appeal to Grice is ultimately uncompelling, the fact that the same bit of discourse may be interpreted either literally or figuratively does lend support to the claim that the difference between simile and literal comparison is interpretive. The second main difference between simile and literal comparison, on Fogelin's view, is that their interpretations differ in how they rank the salience of the features of each of the items compared.[21] Fogelin claims that the salience rankings of the simile are the reverse of the salience rankings of its corresponding literal comparison. To see why this claim is too strong, let us look at one of Fogelin's examples.

Fogelin introduces 'Margaret Thatcher is a bulldozer' as an example of a sentence which most theories of metaphor would designate as false but which his designates as true. He claims that it is true because it is really a way of saying "Margaret Thatcher is *like* a bulldozer" and that sentence is true. The explicitly comparative statement is false if we treat it as a literal comparison since this means allowing "*the standard features* of bulldozers to fix the relevant feature space" (*op. cit.*, p. 87, my emphasis). To do this would violate some key Gricean maxims, and that is our tip-off to treat it as a figurative comparison. When we treat it as a figurative comparison, we reverse the order of salience of features which is established on the literal interpretation.

---

[21] Note that this does not take into account the salience of the features of the context of utterance, which are crucial in taking a Gricean approach to the twice-true case discussed below. In Fogelin's examples, the figurative arises within the sentence, and the context of utterance drops out of the picture. This is one very serious flaw in his appeal to Grice

Fogelin cites what he takes to be a highly salient point in bulldozer feature space, saying that Margaret Thatcher cannot "move huge quantities of dirt in an efficient manner" (*op. cit.*, p. 87). Since she cannot (really) move (real) dirt efficiently (by the standards set by heavy machinery?), uttering this statement flouts the Gricean maxim "do not say what you believe to be false." Bear in mind that bulldozers move more than just dirt efficiently; they move rocks and fallen trees and construction debris and all sorts of things. A more generic description of what bulldozers do is that they push aside whatever is in their path (depending, of course, on the size of the bulldozer and the size of the obstruction). If we keep this more generic feature in mind, then saying that Margaret Thatcher is a bulldozer does not flout the Gricean maxim, for she is a strong woman, capable of quite literally pushing aside what comes in her path. (And metaphorically, she has been quite capable of pushing aside the liberal opposition.) To validate Fogelin's claim that this is a violation of the Gricean maxim, we must amend the maxim to state, "Do not say what you believe to be *false on a literal interpretation*." Unfortunately, the revised maxim is much too restrictive.

Remember that Fogelin claims that in treating a comparison figuratively we reverse the order of feature salience established by a literal interpretation. On a literal interpretation, a highly salient feature of bulldozers is that they are machines. What Fogelin may have had in mind in saying that the order of feature salience is reversed in the move from literal to figurative interpretation is that, once we have called Margaret Thatcher a bulldozer, it does not follow that since bulldozers are machines that we can then call Margaret Thatcher a machine.[22] The inferences sanctioned by a literal interpretation do not follow. But of course, sometimes they do. There is nothing anomalous about saying "Margaret Thatcher is a bulldozer. She's a big grey machine with the power to push aside whatever gets in her way." If we interpret 'machine' literally, we may agree with Fogelin, but if we interpret it figuratively, as we would if we were interpreting 'bulldozer' figuratively, then we must disagree with him.

---

[22] He explicitly offers a parallel argument in discussing 'Winston Churchill is a bulldog'.

The point is that figurative interpretation supports further figurative interpretation, not that we are restricted in what we go on to say when we are interpreting figuratively. If we mix the figurative and the literal in interpreting bulldozer features, then we get the anomalous results. The Margaret Thatcher example shows that we cannot find the distinction between simile and literal comparison in whether we allow the standard features of bulldozers to determine the relevant feature space, as Fogelin contends. This is because the expressions designating those very features may be interpreted either literally or figuratively and so cannot be used to distinguish the literal from the figurative.[23]

II.2 Thesis I (see above, p. 349). Any appeal to the literal meaning of a metaphor is problematic, unless one thinks that metaphors are syntactic types of some sort (and Fogelin does not think this). Fogelin very quickly admits that "in general, of course, sentences of the form 'A is a ϕ' do not have the same literal meaning as sentences of the form 'A is like a ϕ' " (*op. cit.*, p. 29). The problem with Fogelin's first thesis is that he wants to limit it to metaphors but he holds (correctly) that metaphors arise from interpretation and so he cannot really talk about a literally interpreted metaphor. Let me explain.

Apply the first claim to 'Fran is a fox'.[24] Literally interpreted, this means that Fran is a member of a certain class (foxes). The simile 'Fran is like a fox' is *compatible* with Fran's being a fox, but it does not assert that Fran is a fox. We have already seen that on a literal interpretation the 'like' weakens the claim to which it is added. What we interpreters confront is a sentence, a group of sentences, a text, a bit of discourse. Then we make metaphors in our interpreting. *Interpreted literally, there is no metaphor.* 'Fran is a fox' may be interpreted either literally or metaphorically, but it cannot be a literally interpreted metaphor. So, when combined with the view that metaphors are made by interpretation and not just found in texts, Fogelin's own qualifications that this conflation of the meaning of 'A is a ϕ' with the meaning of 'A is like a ϕ' does not hold with just *sentences* serves to undermine the first of the two theses that characterize the comparativist position.

---

[23] When these expressions are interpreted figuratively, and are so interpreted because the initial expression is figuratively interpreted, we then may call the subsequent (figuratively interpreted) expressions "extensions." Metaphors and similes are extendable; literal comparisons are not.

[24] Fogelin uses 'A is a ϕ' to include not only predications of properties but also all forms of metaphor (*op. cit.*, p. 29).

II.3 Thesis II (see above, p. 349). Fogelin takes the second thesis to depend upon the first. He thinks that "since metaphors and similes say the same things literally, their mismatch with context will give rise to the same figurative meaning" (*op. cit.*, pp. 30–1). The second thesis does not fall prey to the argument against the first, since metaphors and similes arise from figurative interpretation and so may be said to have figurative meanings (once interpreted). Since the first thesis is undermined, it would be best to take the second thesis independently. We must ask whether the figurative interpretation of a sentence of the form '*A* is a ϕ' must be the same as the figurative interpretation of '*A* is like a ϕ'. If so, thesis II would be correct.

If correct, thesis II would require: (A) that the semantic feature of 'like' that serves to weaken the assertional commitment of the literally interpreted claim does not operate on a figurative interpretation; (B) that the semantic feature of the 'like' that serves to signal that the claim is comparative when literally interpreted does function when the claim is interpreted figuratively; and, (C) that figurative interpretation cannot help but be comparative. (If true, (C) renders the 'like' eliminable.) The problem for the nonreductive simile theorist is to support (A), (B), and (C). The obvious way to do this is to support (A) and (B) independently, and then use them to support (C). Unfortunately, only (B) is true and then only in some cases. My strategy is to show that both (A) and (C) are false, and to let the simile theorist have (B).

Contrary to (A), 'like' as a weakener of the speaker's assertional commitment does function in figurative interpretation. To see this we must begin by examining the assertional commitments of a metaphor and then examine the assertional commitments of its corresponding simile. Shakespeare's Romeo does not simply say "Juliet is the sun," he says "Ah, what light through yonder window breaks, It is the East, and Juliet is the sun." Most metaphors are extended in their home contexts, and these contexts serve to limit the range of their possible extensions. Shakespeare's context serves to refine the metaphor's assertional commitments and to direct the audience to a limited range of endorsed extensions. In Shakespeare's play, when Romeo says "Juliet is the sun" it is *not* legitimate to extend the metaphor with "She's mostly made of gas." On the other hand, if we take the sentence out of Shakespeare's context and ask "Which extensions are permissible now?" we will find that the whole set of sun-related predicates becomes available. So, if all Romeo says is "Juliet is the sun," then Romeo is committed to licensing all the inferences one can draw from that: that she is 93 million miles away, that she is a celestial body, that she is mostly gas, and so on.[25] Obviously, Romeo has not specifically endorsed or asserted each of these extensions, but he also has not precluded their assertion by others based upon his metaphor. He has licensed the audience to use these extensions without specifically asserting or endorsing any.

Now we must consider the corresponding simile, and again, let us take it out of the Shakespearean context. If Romeo had said instead "Juliet is *like* the sun," and nothing else, he is not necessarily committed to any and all extensions of 'Juliet is like the sun' (as figuratively interpreted), though he is committed to some subset of the extensions. Which subset? We find out only by getting him to continue. If Benvolio were to say "Sure Juliet is like the sun—she's full of noxious gases." Romeo could justifiably reply, "I didn't say that she *is* the sun; I said she's *like* the sun. She is the light and warmth of my world." The disclaimer at the beginning does not preclude Romeo's choosing any particular extensions to endorse. It just signals to his audience that there are some that he will not endorse.[26]

These considerations show that when interpreting a simple unextended metaphor all the possible extensions are open to the audience, but when interpreting a simple unextended simile the use of 'like' suggests a more limited endorsement. In that case, all the extensions are open to the speaker, but not all are open to the audience, and until the simile gets extended, the audience cannot tell which extensions are unavailable. Thus, the 'like' in the simile does weaken the assertional commitment of the figuratively interpreted claim.

The other significant semantic feature of 'like' is that it signals that the sentence is a comparison. This feature must operate on a figurative interpretation if thesis II is correct. ((B) makes this claim for similes; (C) makes the claim for all figurative interpretations.) If we only look at simple identities and sortal predication metaphors, the addition of 'like' does seem to make the resulting sentence comparative. Again, we must ask whether the original metaphor was comparative, just because we can make a comparative claim out of it.

The problems that arise when we turn to substitution metaphors, noun-function metaphors, and verb-function metaphors suggest that (contrary to (C)) the original metaphor need not be seen as comparative at all. In discussing verb-function metaphors, we considered the hyperbolic 'When he was young Sandy Koufax could throw a strawberry through a locomotive'. There we saw that putting a 'like' into the original statement also requires adding an ineliminable quantificational term. We concluded that to understand the metaphor we do not go through these grammatical transformations but instead just ask, "What would Sandy Koufax have to be like for this to be true?" We can get rid of the 'like' altogether by asking instead how he would be, what traits he must have. We are not comparing Sandy Koufax to someone who can throw a strawberry through a locomotive; we are noting that he has a very fast arm and a very keen eye. So, it is not the case that figurative interpretation cannot help but be comparative.

---

[25] For more on this, see my "Extending. The Structure of Metaphor "

[26] In this respect the simile wrenched from context is similar to an extended metaphor within a context, except that the extended metaphor also directs the audience to a set of endorsed extensions while the simple simile does not.

With (A) and (C) undermined, thesis II is undermined, and with it goes the comparativist position as Fogelin has defined it. Although Fogelin takes thesis I and thesis II to define his position, we need not. Fogelin appeals to a simplified Gricean framework to establish an account of figurative interpretation, and in the next section we will examine the strengths and weaknesses of this appeal.

II.4 Figurative interpretation. Fogelin holds that figurativeness depends upon an incongruity between the utterance (interpreted literally) and its context of utterance. Fogelin's distinction between the literal and figurative requires the speaker's "*putting forward a claim figuratively rather than literally*" (*op. cit.*, p. 96) and the audience's interpreting that claim to get a good fit with context. The audience knows the speaker has put the claim forward figuratively if the speaker flouts Gricean conversational maxims. This may seem fine in some cases, but it is undermined by what Ted Cohen[27] has called "twice-true" metaphors.

Consider, for example, the case of the company president who looks out the boardroom window at a sky full of stratus clouds just before meeting with the company's union labor leader. The company president says to her assistant, "There's a storm brewing."[28] There is no incongruity between a literal interpretation of 'storm brewing' and this context, so on Fogelin's view a figurative interpretation would not be supported. But the figurative interpretation of

---

[27] "Notes on Metaphor," *Journal of Aesthetics and Art Criticism*, XXXIV (1976): 249–59

[28] Notice the difficulty of generating the corresponding simile for this metaphor. 'It looks like there's a storm brewing' just seems to be a weakening of the assertional commitment and suggests a more literal interpretation. Note also that the literal interpretation is not necessarily comparative

'storm brewing' is appropriate as well. Lack of fit with context is not the mark of the appropriateness of the figurative, nor is it necessary for the figurative.

We might try to strengthen Fogelin's position by pointing out that it is not simple lack of fit that he has in mind. He says that "with a figurative comparison the speaker flouts, or at least violates, standard [Gricean] conversational rules and thus engages the respondent in the task of making adjustments that will produce good fit" (*op. cit.*, p. 96). Basic to these maxims is adherence to the "cooperative principle," which enjoins us to "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the exchange in which you are engaged."[20] So, the speaker violates, or at least seems to violate, a maxim, such as "Be relevant," "Do not say what you believe to be false," or "Avoid obscurity of expression," and the audience tries to assign a meaning to the utterance such that the speaker has lived up to the cooperative principle.

Let us go back to the twice-true case. Notice that a lot depends on what we take the purpose of the exchange to be, and that in this case, as in most, there is more than one possibility. On the one hand, if we take the company president to be talking about meteorology (perhaps expressing a concern that the weather will make a delayed departure from the office inconvenient), then her gazing at the stratus clouds outside is an important part of the context and the presence of those clouds will help to support the truth of what she says. On the other hand, if we take her to be talking about the impending meeting with the labor representative (perhaps informing the assistant that the meeting will be rough), then other features of the context are more important than the clouds outside the window. And then, since people are complex and often have multiple motivations for what seem to be single actions, it would seem that we could interpret the sentence both ways at once.

In general, it is hard to make a case for the claim that a statement about the weather made while the speaker is looking out at the sky is an act of flouting or even violating conversational maxims. We need to turn our attention away from the speaker for a moment and focus instead on the listener. If the listener were to begin the interpretive process by assuming that the only relevant features of context are those pertaining to the meeting, then he or she may come to wonder why the boss is now commenting on the weather (something low on the scale of relevance). In such a case, the listener might take the utterance to fail to live up to conversational maxims, and so would have to find a way to make the utterance relevant. The listener may decide that the best way to do this is to take the remark as really being about the negotiations. (Notice that even with this interpretive structure governed by the assumption that only what pertains to the meeting is relevant, the assistant could stick with the literal and try to find the relevance more directly by asking, "Do you think the storm will make the others late for the meeting?")

---

[20] Grice, "Logic and Conversation," in A. P. Martinich, ed., *The Philosophy of Language* (New York: Oxford, 1985), p. 162–3.

There are two important features of Fogelin's appropriation of Grice. First, as we have seen, twice-true cases show that the same context can support both literal and figurative interpretations. This raises the question whether the indirect speech-act framework, with its supposed required lack of fit with context, is a useful way to distinguish between the literal and the figurative. Second, the Gricean maxims are maxims governing the cooperative exchange of information. Fogelin's contention that the function of both metaphor and simile is to make comparisons puts emphasis on the cognitive and communicative function of figuratively interpreted discourse. This is a natural corrective to old bad classical theories that treat metaphor as a mere rhetorical flourish, but we must not bind ourselves with correctives.

A restraint on audience interpretation that allows the audience to interpret figuratively only if the speaker flouts conversational maxims is too strict. Although we need a view that emphasizes that figurative interpretation may yield information, we do not want to make the conveyance of the information necessary. In the storm-brewing example, the company president may simply have been trying to pass a few minutes before the meeting with simple pleasantries, and only noticed the figurative significance of what she said once she had said it. Or, she may have recognized and enjoyed the fact that the statement could be well interpreted both literally and figuratively. It would be silly to preclude the figurative interpretation simply because the literal interpretation fits. Even if the exchange of information were the most common generic reason for conversation, this surely would have more to do with us and our activities than with what is figurative about the figurative interpretation of discourse. Besides, conversation is just one form of discourse. It has long been a commonplace (though not necessarily a truism) that in literature and poetry the communication of information takes a backseat to other purposes. For all we know, this may be true in daily conversation as well.

In addition to placing less theoretical weight on the communication of information, we need to emphasize that interpretation and the determination of discursive purpose are interdependent. The aptness of a sort of interpretation and the aptness of a particular interpretation are both a function of the purpose of the discourse. But then, the purpose of the discourse is itself a matter of interpretation.

III. CLAIM (0)

(0) Simile and metaphor play the same linguistic role: they make (or prompt us to make) comparisons.

The simile theory of metaphor is misguided. Behind it lies the assumption that similes and metaphors do the same work and that this work is to make comparisons or to get us to make them. Even if we grant this assumption, it provides no reason to believe that the tropes should be given the same analyses. Both my car and my bicycle take me from home to work, but an analysis of how my bicycle manages the task will not explain how my car does it. The argument from similar effect to same cause just is not valid. The simile theorist ignores the role of the description of the effect in the search for its cause. Both an axe and a saw can divide wood, but an axe cannot saw wood and a saw cannot split wood.

The simile theory confuses the task of explaining what a metaphor means with the task of explaining how metaphor (in general) has meaning. It confuses explanation with explication. It may be of heuristic value to treat a particular metaphor (or even every particular metaphor) as a simile in order to fix its meaning but it does not follow from this that metaphor is just implicit simile. And even if metaphor were implicit simile, since the similes constructed from metaphors cannot be understood wholly literally, the simile theory takes us no closer to the comfort of ready-made semantics. There is thus no reason to grant the simile theorist this basic assumption that metaphor is essentially comparative.

Finally, in assimilating simile to literal comparison, the reductive simile theory obscures what is most important to an account of metaphor. The difference between simile and literal comparison lies in the interpretation: '*A* is like *B*' is a simile when interpreted figuratively and a literal comparison when explication or continuation does not involve any figurative interpretation. An adequate account of metaphor will be an account of what it is to interpret a bit of discourse figuratively.

LYNNE TIRRELL

University of North Carolina

REFERENCE AND PRONOMINAL DESCRIPTIONS*

SINCE the appearance of Saul Kripke's seminal article, "Speaker's Reference and Semantic Reference,"[1] there seems to be a growing consensus that the distinction between attributive and referential uses should be regarded as a phenomenon of pragmatics. In this essay, I shall argue that the consensus is mistaken. Moreover, I believe that the mistake is one with considerable significance. The nature of the distinction has been misassessed because Keith Donnellan[2] and others following him have missed an important generalization. That is, definite descriptions have a number of related *pronominal* uses, and referential use is the use of these pronominal descriptions in direct, singular reference. Pronominal uses, I shall argue, contrast systematically and *as a matter of semantics* with instances of attributive use.

One central strand of Kripke's argument against Donnellan runs as follows. Let us assume, as a first hypothesis, that sentences imbedding definite descriptions uniformly have the truth conditions Bertrand Russell proposed. Accepting this hypothesis does not at all preclude us from going on to distinguish two kinds of case. In one type of case ("the complex case"), a speaker utters ⌜The ϕ is ψ⌝, believing of some object o he has in mind, that it is [uniquely] a ϕ, and meaning to convey to his audience, by means of his utterance, that o is ψ. In "the simple case," the same sentence is produced, but there is no object about which the speaker has the relevant supplementary beliefs and speech-act intentions. Kripke also stresses that, if sentences containing definite descriptions do have Russellian truth conditions, then it is almost inevitable that both simple and complex cases will arise.

These considerations suggest the following view. First, the *literal meaning* of the speaker's words is always given in accordance with Russell's account, and the special "singular" proposition associated with referential use is treated as part of what *the speaker* meant in uttering his sentence. Second, let us say that *the speaker's referent*

---

* Thanks for very helpful comments to Jeffrey C. King, William W. Taschek, and Mark Wilson.

[1] In *Contemporary Perspectives in the Philosophy of Language*, P. French, T. Uehling, Jr., and H Wettstein, eds. (Minneapolis: Minnesota UP, 1979), pp 6–27.

[2] "Reference and Definite Descriptions," *The Philosophical Review*, LXXV (1966): 281–304; see also his "Putting Humpty Dumpty Together Again," *The Philosophical Review*, LXXVII (1968): 203–15.

0022-362X/91/8807/359–87

on a given occasion is the object *o* (if any) about which the speaker meant to convey that it is ψ. It is denied that the fact that *o* is the speaker's referent on an occasion renders *o the semantic referent* of the speaker's token of $\ulcorner$the $\phi\urcorner$. It is not the semantic function of the description token to refer to *o* (if indeed it does) in virtue of *o*'s being the speaker's referent in that context. Within Kripke's framework, the semantic referent of a description is always its unique satisfier. Thus, Kripke, building on Paul Grice,[3] constructs a pragmatic refinement of Russell's theory of definite descriptions, a refined account that purports to offer a better explanation of the data upon which, it has been supposed, a semantical version of the referential-attributive distinction can be grounded. Moreover, Kripke also argues that, if it is assumed that the referential-attributive distinction pertains to the semantics of descriptions, then it is difficult to see how the distinct semantical contributions thus posited are plausibly to be described.[4]

I agree with both of Kripke's conclusions. What I shall attempt to do will be to set out a more complicated range of considerations that do license a semantical distinction between *pronominal* and attributive descriptions. As noted before, Donnellan's referential uses will reappear as a special case of the former. Also, in elaborating these considerations, it will become possible to state a definite proposal concerning the character of the semantical differences between the two general categories.

I

Let us begin with the observation that definite descriptions, like pronouns, may be anaphorically dependent on antecedent terms or phrases. This fact gives rise to a type of ambiguity worthy of attention. The sort of ambiguity I have in mind is easy to illustrate. To begin, consider the statement

(1) Every scientist who was fired from the observatory at Sofia was consoled by someone who knew [him] *the fired scientist* as a youth.

I take it that the italicized description, like the pronoun that could replace it, is a variable bound by the quantifier phrase that fronts the statement. The repeated lexical matter in the descriptor helps here to indicate the tie between the description and the quantifier phrase that binds it. Thus, (1) says

---

[3] For the relevant Grice material, see Part I and essays 15 & 17 in *Studies in the Way of Words* (Cambridge: Harvard, 1989).

[4] Kripke does *not* endorse the pragmatic refinement of Russell's theory briefly sketched above. His main point is that standard considerations in favor of a semantical version of the referential-attributive distinctions are also compatible with the pragmatic alternative, and that the latter seems, in several key respects, methodologically more attractive. He is careful to note in a number of places the choice between these frameworks (and others) is left open by his discussion.

(1) (Every scientist . . . x) (some person y) (y consoled x & y knew x as a youth)

Under certain conditions, however, standard English usage permits the occurrence of a description, similarly bound by the same quantifier phrase, where the lexical indication of the binding is mediated less directly. If a speaker assumes that his audience supposes that

(2) Each scientist who was fired from the observatory at Sofia was a distinguished Bulgarian astronomer.

then, relying on this presupposition, she may utter

(3) Every scientist who was fired from the observatory at Sofia was consoled by someone who knew *the distinguished Bulgarian astronomer* as a youth.

In the imagined context, such an utterance would be little more than a stylistic alternative to (1), i.e., it also states that (1′). What is less clear is how the presupposition (2) of (3) is to be figured into an account of truth conditions. Does (3), under the indicated interpretation, assert the conjunction of (1′) *and* (2), or would the falsity of (2) simply relieve (3) of any truth value whatsoever? Fortunately, present purposes do not require us to resolve this question, but it is an issue that will arise again later.

In any case, the sentence type in (3) clearly has at least one other potential reading. The definite description could be anaphorically independent of the sentence-initial quantifier phrase and occur attributively within the statement. For the time being, let $\ulcorner[\text{the } \phi]_A\urcorner$ mark an attributive occurrence of the description. Then, (3), in such a case, will express

(Every scientist . . . x) (some person y) (y consoled x & y knew [the distinguished Bulgarian astronomer]$_A$ as a youth)

where, I shall assume, the description contributes to the truth conditions of the utterance in the familiar Russellian style.

It should be noted in this connection that the sentence-initial quantifier phrase may be supplanted by a proper name or a demonstrative. Just as we have

(4) Hugo Wexler was consoled by someone who knew *him* as a youth.

we could also have, if the speaker presupposes that Hugo Wexler was a distinguished Bulgarian astronomer, the assertion

(5) Hugo Wexler was consoled by someone who knew *the distinguished Bulgarian astronomer* as a youth.

Modulo the issue about the status of the presupposition of (5), (4) and (5) both say the same thing. Moreover, it seems to me that the pronoun in (4) and the counterpart description in (5) should be treated in a manner parallel to their treatment in (1) and (3)—each pronoun/pronominal description is, in a sense, "bound" by its antecedent. In other words, (1) and (3) are true only if at least one scientist who was fired from the observatory at Sofia satisfies the open sentence

( ) was consoled by someone who knew ( ) as a youth

and, analogously, (4) and (5) are true only if this same open sentence is satisfied by Hugo Wexler. This is to repudiate the idea that, e.g., the pronoun in (4) is a "pronoun of laziness" for the governing name.

Finally, the sentence type (5) exhibits the same ambiguity as the sentence type (3). Clearly, the same description in the same sentence may not "look back" to the name 'Hugo Wexler', and, when it does not, it may have an independent status as an attributive definite description. If so, (5) says

(Someone y) (y consoled Hugo Wexler & y knew [the distinguished Bulgarian astronomer]$_A$ as a youth)

This observation is of interest here since it brings us at least closer to the kind of "referential-attributive" ambiguity that Donnellan has described.

The existence of this kind of ambiguity is undeniable, but it may be felt, with some prima facie justice, that these considerations are so special and so restricted in their application that they represent nothing more than odd exceptions to the more basic rules. But, on the contrary, I believe that these considerations are just the tip of an iceberg: the more fundamental rules are really exemplified here. To substantiate this claim, we need to glance at the intersentential use of pronouns and pronominal descriptions in extended stretches of topic-elaborating discourse.

Some of the key data have become relatively familiar. Here is the beginning of a little story.

(6) Yesterday some man or other left a stick of gum for each of the receptionists. *He* departed unnoticed before *she* discovered *it* on *her* desk.

In a sense that requires explanation, the italicized pronouns in the second statement are introduced by and interpreted in light of the various relevant quantifier phrases in the opening assertion. It has also been noticed that such anaphoric pronouns can often be replaced by suitable descriptions.[5] Thus, in (6), we could have had 'the man', 'the stick of gum', and 'the receptionist' instead of the pronouns that actually occur. Indeed, in some instances, the use of descriptions rather than pronouns becomes obligatory if the intended anaphoric ties are to remain clearly delineated. If (6) had started, 'Some man or other left a stick of gum for some other man at the office', then reduction of unnecessary ambiguity would call for a continuation of the discourse such as

> ***The man who left the stick of gum*** departed unnoticed before ***the man with whom it was left*** discovered it.

It is important to emphasize, as (6) implicitly indicates, that these pronouns and pronominal descriptions can "look back" to universal quantifier phrases. Consider

(7) Any novel Bruno wrote was read by each female critic on the staff.

where this discourse is extended by

(8) She hated it with a passion.

With (7) as its "topic sentence," (8) says

> All of Bruno's novels were hated with a passion by every female critic on the staff.

Similarly, if (7) had been

(7') Any novel Bruno ever wrote was read by at least one female critic on the staff.

then (8), in that context, would have expressed

> Each of Bruno's novels were hated with a passion by at least one female critic on the staff (who read it).

As in our previous example, one or both of the pronouns in (8) might just as well have been a pronominal description instead. A satisfactory account of these terms should provide a unified treatment of the pronouns *and* pronominal descriptions in their relations *both* to existential and to universal quantification.

---

[5] The point is noticed by Gareth Evans in "Pronouns, Quantifiers, and Relative Clauses (I)," in *Collected Papers* (New York: Oxford, 1985), pp. 76–152; and I strongly emphasize the issue in my "Pronouns and Pronominal Descriptions. A New Semantical Category," *Philosophical Studies*, XLV (1984) 1–30

Life and our present exposition would be simple if we could suppose that these pronouns/pronominal descriptions are also variables bound by the quantifier phrases that introduce them into discourse. I take it, however, that it is widely and correctly agreed that this simplest hypothesis is wrong.[6] Unfortunately, the question of a proper semantics for such terms is a topic of considerable controversy in contemporary debate. In what follows, I shall sketch a view that Jeffrey King and I[7] have developed more extensively elsewhere.

It is striking that the pronouns and pronominal descriptions under scrutiny mimic much of the behavior, syntactic and semantic, of the parameters or "arbitrary names" that are standardly deployed in systems of natural deduction.[8] In Benson Mates's[9] system, for example, the following trivial deviations may be constructed:

D_1				D_2		
$\langle 1 \rangle$	(1) $(\exists x)Fx$	P		$\langle 1 \rangle$	(1) $(x)Fx$	P
$\langle 2 \rangle$	(2) $(x)(Fx \rightarrow Gx)$	P		$\langle 2 \rangle$	(2) $(x)(Fx \rightarrow Gx)$	P
$\langle 3 \rangle$	(3) Fa	P		$\langle 1 \rangle$	(3) Fa	UI
$\langle 2 \rangle$	(4) $Fa \rightarrow Ga$	UI		$\langle 1, 2 \rangle$	(4) $Fa \rightarrow Ga$	UI
$\langle 2, 3 \rangle$	(5) Ga	MP		$\langle 1, 2 \rangle$	(5) Ga	MP
$\langle 2, 3 \rangle$	(6) $(\exists x)\,Gx$	EG		$\langle 1, 2 \rangle$	(6) $(x)Gx$	UG
$\langle 1, 2 \rangle$	(7) $(\exists x)\,Gx$	EI				

The term '*a*' in D_1 and D_2, introduced into both derivations as a kind of *schematic singular term*, is an instance of such a parameter. Although it has not been the usual practice to do so, it is possible to assign intuitively correct truth conditions to an arbitrary line containing one or more of these parameters, and, for a well-conceived system of natural deduction, this can be carried out in such a way that each correctly derived formula with parameters is a logical consequence of the premises upon which it rests. The salient fact for present purposes is that the truth conditions of a line with parameters is determined by the (schematic) meaning of the formula *and* by the specific context of the line within the structure of the derivation. Derivational context helps to fix the truth conditions of a line 1 by showing the way in which the parameter in 1 has been introduced (i.e., by EI or UI), by manifesting the relations within the derivation of distinct parameters that occur in 1, and by exhibiting the premises and temporary suppositions from which 1 has been derived. We will notice, if we examine line 5 of D_1 and D_2, respectively, that the formula $\ulcorner Ga \urcorner$ (with the interpretation of $\ulcorner G \urcorner$ held fixed) in D_1 has the force of

---

[6] For arguments against the "bound-variable" view, see the papers cited in fn. 5, and Jaakko Hintikka's "Quantifiers in Natural Languages: Some Logical Problems," in *Game Theoretical Semantics*, Esa Saarinen, ed. (Dordrecht: Reidel, 1978). As noted in the text, I am under the impression that many agree that such uses of pronouns in discourse are not to be treated as bound variables.

[7] "Pronouns, Descriptions, and the Semantics of Discourse," *Philosophical Studies*, LI (1987): 341–63, and my "Pronouns and Pronominal Descriptions."

[8] For an alternative view of the semantics of parameters, see Kit Fine, *Reasoning with Arbitrary Objects* (New York: Blackwell, 1985). Although I am inclined to reject Fine's apparatus of "arbitrary objects," this book is filled with rich and suggestive observations and analyses of the pertinent material. For a comparison of the approach King and I favor to Fine's, cf. King, "Anaphora, Instantial Terms, and Pronouns," *Philosophical Studies*, LXI (1991): 239–65.

[9] *Elementary Logic* (New York: Oxford, 1972, 2nd ed.).

> In each instance (for $\ulcorner a \urcorner$), $\ulcorner Ga \urcorner$ is true.

while the same formula in D_2 conveys

> In at least one *F* instance (for $\ulcorner a \urcorner$), $\ulcorner Ga \urcorner$ is true.

In other words, a formula like $\ulcorner Ga \urcorner$ in these contexts exhibits a systematic ambiguity analogous to the ambiguity of (8) above when it arises out of either (7) or its variant (7′). Much more complicated examples can easily be constructed which involve either parameters in a derivation or pronouns/pronominal descriptions in a discourse, but our simple examples should suffice for the issues before us now.

It is my view, of course, that these pronouns and pronominal descriptions also operate as parameters or schematic singular terms within the confines of a topic-structured conversation, exposition, or narrative. In interpreting such terms, we look to the manner in which they have been introduced as dependent upon antecedent claims and assumptions. When we speak of "a discourse" here, we should include not only the actual utterances, connected by relevance to a larger subject, but also those mutually shared assumptions which the speakers have taken for granted as the conversational framework that guides the series of responsive remarks that they produce. Where we have a pronoun or pronominal description serving as a parameter, we have to trace it back to its spoken or merely tacit source. Often, as in our earlier examples, that source will be a quantifier phrase in some earlier statement or presupposition of the discussion.

These natural-language parameters are variable-like terms that serve in their sentential settings to express appropriate *general* propositions. But they differ from the standard, bound variable of formal logic in at least three important respects. First, they are not literally bound by an antecedent quantifier phrase, but in virtue of context they are, so to speak, semantically unfree. Second, they are descriptionally enriched. If a description ⌜the ϕ⌝ is used as a parameter, the predicables in its description constrain the kind of interpretation the description can receive. In virtue of the fact that English third-person singular pronouns are marked for gender, the same observation holds in a more modest way for them. (I shall consider one significant aspect of this property shortly.) Third, as adumbrated above, extrasentential discourse context determines various features of the proper interpretation of a parameter in a given utterance. That is, (i) the discourse will associate with each parameter a *mode of generality*, either universal or existential as the case may be. Thus, if (8) is uttered as an extension of (7), then 'she' in (8) is associated with existential generality, while 'it' in (8) is universally general. In the same way, the parameter 'a' in the derivations is associated with different modes of generality depending upon whether it has been introduced by UI or EI. (ii) If a statement contains more than one parameter, then the discourse will determine a scope-like *ordering of logical precedence* among them. Looking at (8) again in relation to (7), the knowledge that 'she' is associated with existential generality and 'it' with universal generality does not tell us whether (8) says that each of Bruno's novels was hated by some of the female critics or that some female critic hated all of his novels. So context must and generally does settle this matter as well. Finally, (iii) discourse determines the *domains* over which the pertinent parameters range. If (7) leads into (8), the domain of 'she' in (8) is constituted by the female critics on the staff. The domain of 'he' in the second statement of (7) is made up of the men who left a stick of gum for each receptionist.

In even moderately more complex cases, however, we have to consider the domain of a parameter in a statement *with respect to* an assignment of individual values to other parameters upon which the target parameter depends.[10] Suppose a discourse begins with

(9) (Some x) (some y) Fxy

and the parameters ⌜a⌝ and ⌜b⌝ are introduced as instantiations of the variables 'x' and 'y' in (9). Having done so, the speaker goes on to assert, e.g., ⌜Gab⌝. Intuitively, the domain of ⌜a⌝ is just the set of items that bear F to at least one thing. But "the" domain of ⌜b⌝ should be the set of things to which a bears F. And yet, this condition makes no sense as long as $\ulcorner a \urcorner$ remains a purely schematic term. We obtain a definite domain for $\ulcorner b \urcorner$ only after some object from $\ulcorner a \urcorner$'s domain is assigned to $\ulcorner a \urcorner$ as its referent. In other words, "the" domain of $\ulcorner b \urcorner$ ***depends upon*** a choice of a value for $\ulcorner a \urcorner$, and the concept that we need is: the domain of $\ulcorner b \urcorner$ with respect to an assignment to $\ulcorner a \urcorner$. This allows us to specify correctly that $\ulcorner Gab \urcorner$ is true in the context if

---

[10] For a much more extended discussion of this crucial concept of "(parameter) *dependence*," see the book by Fine and the article by King mentioned in fn. 8. My manner of introducing this concept, briefly sketched below, is somewhat different from theirs.

> (Some x) ($x \in$ the domain of $\ulcorner a \urcorner$) & (some y) ($y \in$ the domain of $\ulcorner b \urcorner$ with respect to x & Gxy)

Therefore, when (7) provides the context for (8), the schema for a domain for 'she' in (8) is $\ulcorner$female critic on the staff who read ( )$\urcorner$. A particular domain is defined when one of Bruno's novels is taken as the value for $\ulcorner( )\urcorner$.[11]

When all of these features—mode of generality, ordering of logical precedence, and independent and dependent domains—have been adequately aligned by the discourse, then a parameterized statement assumes definite and definable truth conditions.[12] I should mention that there is nothing unprecedented about the basic mechanisms at work in this affair. The sentence, 'Elephants fly' is, as it stands, ambiguous or indeterminate between (at least) 'All elephants fly' and 'Some elephants fly'. On a given occasion, context, including the speaker's relevant intentions, will decide the wanted mode of generality. Similarly, the sentence 'Every man loves some woman' is also notoriously ambiguous, and contextual factors must indicate the order of logical precedence between the pair of quantifier phrases. Finally, when I say, "Every napkin is frayed," it will be context that further restricts the range of quantification to be, e.g., napkins in the house, napkins on this table, or whatever. In the case of the parameters, it is simply that all of these features are fixed simultaneously, and the contextual factors that effect the triple determination tend to be structurally more complex. Both the nature and the structural complexity of these discourse contributions are more perspicuously exhibited in the regimented environment of natural deduction systems.

---

[11] These brief remarks are meant only to give an impression of the special semantical complexity that attaches to parameters, but to do so in a way that keeps in focus their relation to the simple and more familiar machinery of quantifiers and the variables they bind.

[12] For a developed semantics for parameters in a system of natural deduction elaborated along the lines I favor here, see King's "Pronouns, Descriptions, and the Semantics of Discourse," and his "Anaphora, Instantial Terms, and Pronouns."

These remarks on pronouns/pronominal descriptions qua parameters are too brief—the topic is a rich and subtle one. They should be enough, however, to extend the central line of thought I was developing earlier. For even if pronominal descriptions are used sparsely as bound variables, their use as parameters that express generality is certainly widespread. Moreover, this use generates new instances of the sort of ambiguity that was illustrated when bound pronominal descriptions were on the stage. Modifying a previous example, we have

(10) (a) At least one } scientist at the observatory at Sofia
(b) Every } was fired.

where either of these is continued by

The fired scientist was consoled by someone who knew him as a youth.

But now, let us imagine once more that the speaker presupposes (2), i.e.,

Every scientist who was fired from the observatory at Sofia was a distinguished Bulgarian astronomer.

If so, then the elaboration of (10a) or (10b) could take the form

(11) The distinguished Bulgarian astronomer was consoled by someone who knew him as a youth.

With the pronominal description read as a parameter that is governed by the presupposition (2), (11) says, in relation to (10a),

(At least one scientist . . . *x*) (some person *y*) (*y* consoled *x* & *y* knew *x* as a youth)

or in relation to (10b),

(Every scientist . . . *x*) (some person *y*) (*y* consoled *x* & *y* knew *x* as a youth)

In contrast with both of these readings, however, there is another in which the pronominal description is anaphorically independent of (10) and occurs attributively in (11). On this reading, (11) states,

(Some person *y*) (*y* consoled [the distinguished Bulgarian astronomer]$_A$ & *y* knew him as a youth)

This type of ambiguity is, it seems to me, quite commonly exemplified.

Given this parallel between pronominal descriptions and third-person pronouns, and given that the pronouns have a ubiquitous *deictic* or *indexical* role, one would expect that pronominal descriptions might also share this role. And, of course, it seems as if they do. When a speaker gestures toward Hugo Wexler in conversation with another person and says,

> The distinguished Bulgarian astronomer is being consoled by someone who knew him as a youth.

then we have an apparent instance of a pronominal description used deictically.

It is precisely the semantic status of such utterances, however, that has been the subject of dispute. Let

(12) The $\phi_{\downarrow}$ is ψ

represent a deictic employment of the description, and let *o* be the object that is contextually indicated as the subject to which the predicate is purportedly ascribed. The parallel with pronouns suggests that the deictic character of the description ***is semantically relevant*** to the strict and literal meaning of the utterance that contains it. That is, the semantic relevance hypothesis (SRH) claims that the strict and literal meaning of (12) on the relevant occasion is such that (12), in that context, expresses the Russellian singular proposition

of *o*, that it is ψ

while depending upon the singular presupposition

of *o*, that it is also ϕ

There is an issue here, as there has been in previous cases, about how the presupposition is to be counted in relation to the strict and literal meaning of the utterance, and in particular, about how the presupposition acts upon the conditions of truth and falsity. For myself, it seems that if *o* is not ϕ, then what is strictly and literally said by the words should be assessed as untrue, but I am not certain whether the untruth is simply false or is defective in relation to truth in some other way. Nevertheless, however this question may be decided, the world-to-world truth conditions of (12) will differ, according to (SRH), from those of an utterance of ⌜The ϕ is ψ⌝ in which the description is attributive. Obviously, deictic uses of pronominal descriptions correspond closely to the occurrences that Donnellan called "referential uses of definite descriptions."

(SRH) seems immensely natural from within the larger framework I have sketched. In fact, it is plausible to maintain that deictic uses of pronouns and pronominal descriptions constitute a limiting case, albeit a central limiting case, of their alternative uses as variables and parameters. Suppose that discourse context can assign to a parameter in a statement—not a mode of generality and domain—but a particular object, an object that is thereby assigned to the parameter as the subject of the conjoined predicate. Indeed, this is exactly what seems to take place if we consider an analogue of the discourse (10) + (11) in which the initial quantifier phrase is replaced by a proper name, e.g.,

> Hugo Wexler was fired from the observatory. The distinguished Bulgarian astronomer [or: He] was consoled by someone who knew him as a youth.

I take it that the second statement, in virtue of the intended connection between the name and the pronominal, says *of* Hugo Wexler that he was consoled . . . , etc. Therefore, it is difficult to see why sundry nonlinguistic features of the speech context should not be able to achieve the same result. A demonstrative gesture accompanied by a plain intention to be speaking of the entity demonstrated could yield a similar assignment to the pronominal parameter and so could other less explicit, more elaborated facets of the speaker's discourse circumstances.

In many systems of natural deduction, the same set of terms is used, with a studied and harmless ambiguity, both as parameters and as individual constants. In any particular derivation, such a term will function in one way or the other, but, from derivation to derivation, the function of a given term may well change. Why should it not be the case that English, with its pronouns and descriptions, incorporates terms that behave similarly? Why should English not have expressions that appear as schematic singular terms in one discourse and then pop up in others as genuine singular terms referring to specific objects? Why not, indeed? I believe that this is precisely the case.

We still have not settled on the truth of (SRH), however. Despite its appeal, the fact remains that Kripke has produced powerful arguments to undermine its various attractions. Therefore, in the next two sections we turn to an investigation of those arguments and of some significant related questions that they raise.

II

In "Speaker's Reference and Semantic Reference," Kripke several times underscores the methodological principle that a "unitary" semantical account of definite descriptions is preferable to any theory that postulates an intuitively dubious semantical ambiguity. The principle is surely correct, and Kripke is right to emphasize its importance. But, from our present point of view, we see that the principle points inconclusively in different directions. We have reminded ourselves that no unitary account in Russell's mode will accommodate all of the interesting, systematic uses of descriptions: the uses of descriptions as bound variables and parameters cannot be so accommodated. Any overview of the data must recognize some broad distinction between descriptions used attributively and descriptions used pronominally.

A defender of Russell will contend, however, that the pragmatic refinement of Russell's theory offers a unitary account of all nonanaphoric uses of description. And, it will be added, this is all that the theory could reasonably be expected to cover. On the other hand, (SRH) contests even this more restricted claim. Deictic uses of pronominal descriptions should be recognized as distinct from the attributive uses, and the deictic use forms a part of a unitary theory of pronominal descriptions. Or, putting the matter in a different way, according to (SRH) the deictic uses are themselves, in a sense, anaphoric. Unlike the standard examples of anaphora, they are anaphorically dependent upon various nonlinguistic contextual determiners of singular reference which are present in or available from the utterer's speech-act situation. Hence, both theories propose a kind of unification of the pertinent phenomena. The difficult question is: "Which form of unification is correct?"

The natural place to look for evidence to decide this question is the plentiful domain of more or less obviously *incomplete* descriptions. Someone says, "The baby is crying" or "The table is covered with books," and it is not a part of what has thereby been said that there is one and only one baby or one and only one table. On the face of the matter, many such statements *seem* to involve deictic, pronominal descriptions, and this looks favorable to (SRH). Nevertheless, it is the standard reply of the Russellian that Russell's analysis is not always to be applied naively to the bare words actually emitted by the speaker. Rather, it is supposed that the speech context will somehow supply a restrictive condition upon the truncated description, a restriction that makes the implication of uniqueness something the speaker could plausibly have meant.[13] Thus, both

---

[13] I should note that Kripke himself sets the issue of incomplete descriptions to one side in his essay, and he does not endorse the emendation of Russell's ideas I shall be discussing. He states: "Contrary to the Russell picture, I doubt that such [incomplete] descriptions can always be regarded as elliptical with some uniquely specifying conditions added. And it may be the case that a true picture will resemble Donnellan's in important respects" (*op. cit.*, pp 6–7). In what follows, I shall argue that Kripke's doubts are well-founded.

(SRH) and Russell's account, emended to cover patently incomplete descriptions, invoke a regular role for context in determining what is literally said by "referential" tokens of $\ulcorner$The ϕ is $\psi\urcorner$, but the character of what is so determined differs substantially between the two proposals.[14]

Up to a point, contextual qualification of some incomplete descriptions can and should be countenanced. For instance, speakers often leave various obvious indexical specifications unstated but easy to supply. When Donnellan's party guest utters "the man who is drinking champagne," it will be clear that he means "the man who is *now* drinking champagne *here*" or something of the sort. In particular, there are many occasions on which a description used attributively will be patently incomplete. A detective, contemplating the horribly mutilated corpse, says "The murderer is insane," but context makes it plain that he is to be understood as asserting "The murderer *of Smith* [or: The murderer of him] is insane."[15] These commonplace styles of supplementation, however, will not generally suffice to yield a uniquely satisfied descriptor. Speaking of a certain piece of furniture in an antique store full of tables, someone states, "The table is covered with books." Here also the Russellian requires that some suitable emendation be recoverable from context, and, at this juncture, it seems to me, the idea is protected only by a convenient fogginess about how context is to do its job.

Presumably, the usual thought has been that there will be a particular descriptive restriction that the speaker has in mind and that he intends to qualify his audible definite description. Now, this vague thought seems to me to be almost certainly wrong, but it is one that admits of so many possible variations that I shall simply attempt to explain why the prospects of any of its versions look quite bleak. The upshot will be that this doctrine is exposed to many of the problems that afflict the similar view that the semantic referent of a proper name token is determined by an associated description.

---

[14] I do not have space to discuss views, like Kent Bach's, that radically reject the need for *any* contextual supplementation of the literal meaning or "content" of utterances that contain incomplete definite descriptions. See his *Thought and Reference* (New York: Oxford, 1988), pp 91–130. On such a view, what has been literally expressed by such an utterance will be, in general, simply false. On the other hand, what the speaker thereby means to convey may well be true. For a critical discussion of Bach on this issue, see William W. Taschek's review of the book in this JOURNAL, LXXXVII, 1 (January 1990). 38–45. I shall note only that it seems highly implausible that, when the proud father says to his wife, "The baby is prettier than any other babies I've ever seen," his words, on that occasion, *literally express a contradiction*.

[15] Scott Soames emphasizes the importance of these considerations in "Incomplete Definite Descriptions," *Notre Dame Journal of Formal Logic*, XXVII (1986): 349–75.

First, there are cases in which a token description intuitively refers to a particular object even though the description is not uniquely satisfied and the speaker has *no resources at all* for qualifying his words adequately. Admittedly such cases tend to have a somewhat *recherché* character, but they nonetheless exist. Suppose that Jean Valjean, as a very old man, has only the spottiest memory of his life up to a year before the events now to be described. One day while walking through Paris he catches sight of a face that fills him with fear and foreboding. When he sees the face he cannot say what it is that he so dreads, and he continues through the day brooding over his troubled state of mind. By later that evening two things have developed. He has totally forgotten that it was the sighting of the face that gave rise to his continued emotional disturbance and, at the same time, he comes to be seized with the conviction that, as he expresses it, "My old enemy is back in Paris." In fact, it is this thought that he now takes to give the object of his fear.

Although Valjean insists that there is a definite person of whom he is thinking, he admits that he has had a number of enemies in the past and he cannot produce any further description to single out the person in question. Finally, imagine that we know, although Valjean does not, facts such as these. The original cause of Valjean's frightened response was the fact that the face he saw bore a definite resemblance to the face of his old enemy Javert. Further, only an intimation, conscious or unconscious, of Javert's presence could generate in Valjean this specific syndrome of intense emotional disturbance. Finally, if Javert, as Valjean had known him years ago, could be directly presented to Valjean now, then Valjean would realize straight off that *this person* is the old enemy he means. In other words, despite the fact that Valjean is unable to summon to mind any helpful representation of Javert, he would be able to recognize the Javert of his past as the man he has been thinking of all along. As facts of this kind are assembled, it becomes hard to deny that Valjean's use of 'my old enemy' does refer to Javert even though nothing like a uniquely identifying description has been provided.

Perhaps it may seem that Valjean has access at least to the description 'the old enemy of mine who is the cause of my present fear'. Let us add to our story, however, one of those amazing coincidences typical of a 19th century novel like *Les Misérables*. The man whose face reminds Valjean unconsciously of Javert is also a former enemy of Valjean's, though Valjean does not register that further fact in any way. Thus, the emended description does not identify Javert uniquely. No doubt there are various other emendations that could be tried, but I think that this last attempt is typical. If the expanded description is not question-begging, a simple alteration of the story will render it improper.

Second, especially in situations where the object referred to is well-known to the speaker and to his audience, and this fact is mutually known, there will be a wide range of qualifications which the speaker is in a position to offer but no one of which is something the speaker clearly had in mind as meant to qualify his words. Also, some of these potential qualifications may yield a uniquely denoting description but many others will not, or, at least, the qualified descriptions will not uniquely denote the intuitive referent. The Russellian must therefore find considerations that *determine* which of the available qualifications is to play the role, as he sees it, of determining the relevant designation. To this challenge there is, I suspect, no satisfactory solution.

If one takes the qualification the speaker would have given at the time of utterance if he had been asked ⌜Which ϕ?⌝, then—assuming (wrongly) that the counterfactual here is, in general, well-defined—there will be examples in which the emended description, but not the original, misdescribes the ϕ in question. A man says to his wife, "The girl ought to be punished," and he would naturally and spontaneously fill this in as 'the girl who is our daughter'. But, of course, he can be mistaken about his relationship to the girl. Even if he is allowed to weaken his favored elucidation to 'the girl whom I believe to be our daughter', this description will be improper if the wife has managed a perverse domestic juggling of identical twin girls. There are many cases in which it is clear which object is referred to and what facts determine that object to be the referent, but where it is totally obscure what further identificatory phrase is supposed to qualify the uttered description in the way the Russellian demands. The most obvious choices will sometimes yield intuitively wrong results, and it is difficult to surmise what plausible but less obvious criterion of "qualification determination" could be advanced.[16]

---

[16] I also shall not explicitly discuss alternative Russellian approaches that do not treat incomplete descriptions as essentially elliptical but do require context to provide a semantically relevant contextual supplementation It has been held that speech context can supply a *domain* of individuals D such that the relevant utterance is true just in case there is one and only one object that is ϕ in D and it is ψ. As a rule, D is not determined through an implicit supplementation of the patently incomplete description. Versions of this idea may be found in Jon Barwise and John Perry, *Situations and Attitudes* (Cambridge: MIT, 1983); and Jaakko Hintikka and Jack Kulas, *Anaphora and Definite Descriptions* (Dordrecht· Rei-

The package that consists of the pragmatic extension of Russell's theory combined with a Russellian treatment of incomplete descriptions faces a related, serious difficulty. Each of the sentences

(12a) He who murdered Smith is insane.
(12b) This man who murdered Smith is insane.
(12c) That man who murdered Smith is insane.

contains subject terms that have attributive or attributive-like uses. The sentences, in appropriate circumstances, may be used so that they are strictly equivalent to—in fact, so that they are stylistic variants of—the sentence

(14) The man who murdered Smith is insane.

where this description occurs attributively. Such a use of (12b) and (12c) will perhaps be most natural in a situation when the discussion has reverted to, for example, previous speculations about who may have murdered Smith. Now, *if* the Russellian proposals concerning incomplete descriptions were to succeed, then it is hard to see why the same proposals should not work for demonstratives as well. Let

(14a) He$_\downarrow$ (she$_\downarrow$, it$_\downarrow$) is ψ
(14b) This $\phi_\downarrow$ is ψ
(14c) That $\phi_\downarrow$ is ψ

be statements containing demonstratives that have been uttered in a context *C*, and let *o* be the item referred to in *C* by means of the relevant demonstrative. One might then suppose that *o* is the semantic referent in *C* of the demonstrative and that the imbedding utterance strictly and literally expresses the proposition

(15) of *o*, that it is ψ

A new application of the "package" under scrutiny will inform us, however, that this is wrong. For these proposals will tell us that the subject terms in deictic tokens of (14) are really incomplete versions of the expressions 'He who . . .', ⌜This ϕ who . . .⌝, and ⌜That ϕ who . . .⌝, where a suitable completion of the material elided will be indicated by the speech context. If ⌜θ'd⌝ represents that completion, then the strict and literal meanings of (14) will be given by

---

del, 1987). I believe, however, that many of the issues I raise in the text against the "supplementation" view are easily extended to generate problems for these alternatives. For an incisive critical discussion of the specifics of the Barwise-Perry treatment of incomplete descriptions, see Soames, "Incomplete Definite Descriptions."

He who θ'd is ψ
This ϕ who θ'd is ψ, etc.

with the subject terms interpreted attributively in each case. On this view, the Russellian singular proposition (15) is to be regarded as a content that the speaker, through his utterance, merely implicates.

I assume that there will be wide agreement that this is an overzealous application of the pragmatic machinery Kripke invokes. Ignoring the difficult case of proper names, this application of the machinery will leave, at best, only 'I' and some purely indexical uses of terms like 'here' and 'now' within the category of terms that refer directly to their semantic referents. Although Russell himself might have been happy with such a conclusion, the result is, I take it, quite unsatisfactory. Therefore, the unrestrained proponent of the Russellian package is faced with a dilemma. Either he gives up Kripke's pragmatic revision of Russell from the beginning, or he rejects the Russellian doctoring of incomplete descriptions. If the second option is adopted then he has no account, alternative to (SRH), of "deictic" incomplete descriptions. And, finally, since the incompleteness of a description cannot by itself be crucial to its *semantic* status, he had better acquiesce in a semantic account of deictic uses of pronominal descriptions that are, as it happens, satisfied uniquely.

Thus far, all we have assumed about attributive descriptions is that statements containing them have Russellian truth conditions. But Russell's own investigations of definite and indefinite descriptions suggest a further thesis: the thesis that ⌜the ϕ⌝, when it occurs attributively, is a quantifier phrase of English.[17] If ⌜The ϕ_A is ψ⌝ has the truth conditions of ⌜Something or other which is uniquely ϕ is ψ⌝, then it seems a natural and modest hypothesis that ⌜the ϕ_A⌝ just *is* a quantifier phrase that means (roughly) ⌜something or other which is uniquely ϕ⌝. Therefore, ⌜the ϕ⌝, synonymous with ⌜the *only* ϕ⌝, is to be classified with (nonanaphoric and nondeictic uses of) the following: ⌜the ϕs⌝, ⌜the few ϕs⌝, ⌜the several ϕs⌝, ⌜the n ϕs⌝, and ⌜the many ϕs⌝. These all are English quantifier phrases, the first in the singular, the rest in the plural. Let us adopt the convention of writing $\ulcorner \ldots Q\phi \ldots \urcorner$ as $\ulcorner (Qx{:}\ \phi)(\ldots x \ldots)\urcorner$. Now, our present proposal, combined with our previous discussion, tells us, in effect, that, if $\ulcorner$The ϕ is $\psi\urcorner$ contains an attributive description, it is properly represented by

---

[17] I shall not attempt here any positive defense of the view that descriptions, used attributively, are quantifier phrases. Among recent writers on the subject, the view is endorsed both in Evans, and Stephen Neale, *Descriptions* (Cambridge: MIT, 1990). For Evans, the descriptions are *binary* quantifiers; for Neale, they are *restricted* quantifiers. For present purposes, *any* view of the semantics of definite descriptions will do which (i) yields the result that they engender Russellian truth-conditions and (ii) do not belong in the category of definite singular terms. For more on the issues raised by this second condition, see the conclusion of the following section

$$(\text{The } x{:}\ \phi)\ \psi x$$

while, if the contained description is pronominal and deictic, the statement has the form of

$$\psi(\text{the } \phi_{\downarrow})$$

where $\ulcorner$the $\phi_{\downarrow}\urcorner$, as before, is an indexical singular term. For the remainder of this paper, I shall take the phrase 'the semantical version of the referential-attributive distinction' to refer to this hypothesis.

This or any similar proposal remains suspect, however, in the face of intuitions that the meaning of the definite article is univocal in both of the two uses. Given the proposal above, $\ulcorner$the $\phi\urcorner$, when it serves as a quantifier phrase, carries the uniqueness condition, but when it is a pronominal description used deictically it does not. Conversely, a deictic pronominal description purports to refer rigidly to a particular item, but the quantifier phrase does not. In the light of contrasts like these, one would expect there to be a reasonably sharp linguistic intuition that the meaning of the words, e.g., "the murderer of Smith" is different between the two uses. But no such *definite* intuition seems to be forthcoming. Are we not in danger of postulating a semantic ambiguity that our basic semantic competence in the use of descriptions does not plausibly sustain?

I think that these remarks express at least some of the concerns behind a methodological consideration that Kripke brings to bear against a semantic account of the referential and the attributive. He states,

> . . 'Bank' is ambiguous, we would expect the ambiguity to be disambiguated by separate and unrelated words in some languages. Why should the two separate senses be reproduced in languages unrelated to English? First, then, we can consult our linguistic intuitions, independently or any empirical investigation. Would we be surprised to find languages that used two separate words for the two alleged senses of a given word? If so, then, to that extent our linguistic intuitions are really intuitions of a unitary concept, rather than of a word that expresses two distinct and unrelated senses (*op. cit.*, p. 19).

A little later, he continues,

> . . . I at least would find it quite surprising to learn that, say, the Eskimo, used two separate words "the" and "ze", for the attributive and referential uses. To the extent that I have this intuition, to that extent I think of "the" as a unitary concept (*op. cit.*, p. 19).

How can we explain the apparent upshot of this intuitive test?

I think that our continuing analogy between pronouns and pronominal descriptions is quite instructive in relation to these last questions. Consider the following examples:

> If any man marries too young, *he* will regret it.
> Some cowboy stopped at the camp. *He* rode a horse.
> Every cowboy worked the range. *He* rode a horse.
> He_1 rode a horse.
> *He* who murdered Smith is insane.

and, exploiting a common idiom,

> George Eliot was not a *he*.

Suppose one were to ask whether the word 'he' in these statements does or does not have the same lexical meaning. It is my impression that the query is likely to lead to confusion and, perhaps to an answer like, 'Well, it does and it doesn't'. The problem arises, it seems to me, because our intuitive concept of "lexical meaning" is too crude to do well in such a case, and it is this crudity which confuses the linguistic intuitions we try to describe using the defective concept.

Up to a point at least, it is easy to make out where the difficulty lies. On the one hand, the semantic function of the pronoun is different in all or most of these examples. In that respect, the "meaning" of the pronoun can be said to vary from one statement to the other. On the other hand, the descriptive conditions or constraints upon the correct "application" of the pronoun remain constant through all of these occurrences. That is, the term "applies" correctly to a potential subject of application only if that entity is a male. No doubt, the word 'applies' in this characterization is either ambiguous or it is an umbrella term that covers all of the pertinent semantic roles. Still, how it is to be understood in each case is clear enough. Therefore, in this respect, we might well be led to say that 'he' has the same lexical meaning in each of our last examples. (We would not expect to find a separate dictionary entry for all or even most of these uses.)

In considering these statements, we see why a standard example like 'bank' gives us a bad model for grasping the "ambiguities" before us now. 'Bank', used in relation to financial institutions and sides of rivers, retains the same kind of semantical role. In both of its senses, the word 'bank' is a predicable—more specifically, a general term. The two senses arise because the term is conventionally associated with two sets of applicability conditions. This is the reverse of the situation with regard to the pronoun. The word 'he' carries its minimal conditions of applicability along as it systematically shifts the semantical function it performs in the utterances in which it appears.

It is my view that we ought to look at definite descriptions in *their* various semantical roles, in the same way. A given description does not change the conditions of correct application which it semantically expresses in the descriptor even when it does change from one pronominal function to another or from its pronominal status to its attributive use as a quantifier phrase. The "descriptive sense" (as we might put it) of, e.g., 'Smith's murderer' is just the same no matter whether the term occurs as a quantifier or as a pronominal description of one stripe or another. In this way, we can accept and explain the intuition that the descriptive phrase expresses "a unitary concept" while maintaining, at the same time, that very different semantical roles should be distinguished.

I have not attempted a direct answer to Kripke's question about whether we would expect pronominal and attributive descriptions to translate into distinct linguistic forms in languages other than English, and I have not done so for a reason that he himself supplies. In a footnote, he briefly considers some ways in which his intuitive test "needs further explanation and refinement." He says,

> . . . a word may have different senses that are *obviously related* [the italics are mine]. One sense may be metaphorical for another (though in that case, it may not be a separate sense, but simply a common metaphor.) "Statistics" may mean both statistical data and the science of evaluating such data and the like. The more we can explain relations among senses, and the more 'natural' and 'inevitable' the relationship, the more we will expect the different senses to be preserved in a wide variety of languages (*op. cit.*, p. 26, fn. 29).

I have tried to indicate why I believe that gendered pronouns and definite descriptions in English provide beautiful, distinctive, and fairly complex examples of the situation Kripke describes in this passage. The terms in question have a family of semantical functions that are closely and naturally related to one another, and, as a consequence, our expectations about how other languages will distribute these functions within their lexicons are, on the whole, pretty dim. We should not allow the nagging feeling that *somehow* the "lexical meaning" of descriptions is "the same" dissuade us from the conclusions we have drawn.

III

In this section, we turn first to an important and influential objection that Kripke raises against any semantical competitor to the pragmatic refinement of Russell's theory. A fairly careful examination of this objection will help considerably to clarify some of the major points that divide the semantical version of the referential-attributive distinction from its quasi-Gricean alternative.

Kripke begins by pointing out that there are simple and complex cases involving *proper names*. An instance of a complex case is described in the following passage:

> Two people see Smith in the distance and mistake him for Jones They have a brief colloquy: 'What is Jones doing?', 'Raking leaves.' 'Jones' in the common language of both is a name of Jones, it *never* names Smith. Yet, in some sense, on this occasion, clearly both participants in the dialogue have referred to Smith, and the second participant has said something true about the man referred to if and only if Smith is raking leaves. . . . How can we account for this? (*op. cit.*, p. 14).

In the corresponding simple case, there is no man seen in the distance. The speaker says, simply using 'Jones' as a name of Jones, 'Jones is raking leaves'. This utterance might express, for instance, an educated guess about what Jones, wherever he is, is doing at that time. Considering the two cases, Kripke points out that it lacks all appeal to account for the differences between the cases by postulating a semantic ambiguity either in the name or in the sentence that contains it. In both situations, the semantic reference of the name is Jones and the sentential utterances say, strictly and literally, *of him* that he is raking leaves. In the complex case, because of the speaker's conflation of Smith with Jones, it is also true that the speaker intended *of Smith* that the name token refer to him. But, although *the speaker* was, in a sense, referring to Smith, the name, even on that occasion, did not. To repeat, a semantic ambiguity is out of the question in these cases, and it looks as if the familiar pragmatic machinery will handle the differences perspicuously.

In fact, Kripke uses his pair of name examples to state two difficulties for any proposed semantical version of the referential-attributive distinction. The first difficulty has already been discussed. It is based upon the following subtheses. (i) The pragmatic account is independently required to deal with simple and complex cases involving proper names, and (ii) this same account is also adequate to explicate the phenomena involved in "referential uses" of descriptions. From these, it is inferred that the pragmatic account is to be preferred. Now, (i) is true, of course, and this establishes that the existence of simple and complex cases with definite descriptions does not support a semantical distinction for descriptions. What does support such a distinction, from my perspective, is the more complicated set of considerations described in sections I and II. Correspondingly, these same considerations make the case against thesis (ii). It is not plausible that all deictic uses of pronominal descriptions can be treated in a satisfactory way by means of the pragmatic apparatus. Therefore, what Kripke here calls "the unitary account" is not as broadly adequate as these last remarks of his suggest.

But Kripke goes on to formulate a second puzzle. "Further, the resemblance between the case of descriptions and that of proper names (where presumably no one would be tempted to postulate an ambiguity) is so close that any attempt to explain the cases differently is automatically suspect" (*op. cit.*, p. 20). This passage can be construed as offering a challenge to the semantic version of the referential-attributive distinction. Since the phenomena in the pair of name cases are so closely analogous to the phenomena in the counterpart attributive and referential cases, it is extremely difficult to understand how it is that there is no semantic ambiguity in the name cases while, in the description cases, such an ambiguity supposedly exists. The challenge to the advocate of a semantic distinction for descriptions is to explain how this can possibly be.

Oddly enough, a crucial hint as to how this challenge should be met is found in another of Kripke's footnotes.

> There is one significant difference between the case of proper names and that of definite descriptions. If someone uses "Jones" to refer to Smith, he has ***misidentified*** Smith as Jones, taken Smith for someone else. To some extent I ***did*** think that *Jones* was raking the leaves. (I assume that "Jones" is already in his idiolect as a name of Jones. If I am introduced to an impostor and am told, "This man is none other than Albert Einstein," if I am fooled I will have ***taken*** him, falsely, to be Einstein. Someone else, who has never heard of Einstein before, may merely be mistaken as to the impostor's name.) On the other hand, if I think that someone is "her husband" and so refer to him, I need not at all have confused two people. I merely think that one person possesses a property—that of being married to her—that in fact he lacks. The real husband is irrelevant (*op. cit.*, p. 25, fn. 26).

The delicate issue that this passage raises but does not attempt to answer is: "In what sense is the actual husband 'irrelevant' to the statement?"

The answer that I have proposed, in effect, proceeds along the following lines. Consider the assertion

(16) Brown is her husband.

In one obvious and perfectly standard use, (16) means

Brown is a man who is (uniquely) married to her.

That is, the predicate 'is her husband' simply ascribes the indicated complex property or condition to Brown. In relation to the predicate, only this property is semantically (and pragmatically) relevant. If we imagine, for the sake of illustration, that (16) is false and that the lady is actually married to Smith, then (16) does not express, strictly and literally, the singular proposition

of Smith, that Brown is (identical with) *him*

Smith, the actual husband, is semantically (and pragmatically) irrelevant to the meaning that the predicate, on that occasion, expresses. Now, once this is allowed, there is no reason not to recognize that definite descriptions in other grammatical positions may function in the same way. Thus, there is a natural (attributive) reading of

(17) Her husband is kind to her.

which says (roughly)

A man [some man or other] who is (uniquely) married to her is kind to her.

In this utterance also, 'her husband' functions as a kind of quantifier phrase (i.e., 'The *x:* husband of her'), and the utterance does not even purport to say *of* Smith that he is kind to the woman. That is, the actual husband is once more—and in the very same way—semantically irrelevant. For the advocate of semantically ambiguous descriptions, this is the crucial difference that is *not* reflected in simple and complex cases with proper names. The words 'Jones is raking leaves', in either case, literally say of Jones that he is raking leaves—they literally express a singular proposition of which Jones is a constituent. But, by contrast, the semantical version of the referential-attributive distinction explicitly endorses the point suggested by Kripke's footnote. If the description in (17) is attributive, then the proposition literally expressed by the words is a *general* proposition (at least in relation to what the full description contributes).[18] It is only when the description is pronominal and used deictically that a genuine singular proposition is expressed. Therefore, since the purported contrast between the attributive and the referential, properly understood, is of a different character from the contrast between simple and complex uses of names, the absence of a semantical ambiguity in the names does not impugn the coherence and integrity of the semantical distinction for descriptions. To repeat a previous disclaimer, however, the evidence that this distinction is actually exemplified in our use of descriptions must depend on facts other than the mere existence of simple and complex cases that feature definite descriptions.

These contentions will be substantially clarified if we see that Kripke's pragmatic apparatus does indeed have a very general applicability. In particular, it has an important role to play *even within* a framework that adopts the semantical version of the referential-attributive distinction. The familiar pragmatic discriminations are required in cases where a description occurs attributively *and* in cases where a pronominal description is used deictically.

Here, first of all, is an "attributive" example typical of what I have in mind. Greg expresses to Brian his desire to earn $50 so that he will be able to buy a new bike, and he wonders how he might get the money. Brian recalls that the annual Cub Scout lightbulb sale is just underway and that the most successful Scout salesman wins a $50 prize. He also knows, as does Greg, that Greg has been the winner for two years running and that, given his superlative sales talents, Greg is sure to win again this year. So Brian, meaning to remind Greg of all this, says

> The Scout who sells the most lightbulbs this year will win $50.

Brian's use of the definite description is plainly attributive and expresses a thought about whomever sells the most bulbs. It is *not* a pronominal description making reference to Greg. In this context, however, Brian has intended to suggest to Greg that he (Greg) will shortly win a $50 prize. But this thought, the chief thought behind Brian's remark, is not something literally expressed by the words even on this occasion. It is merely a thought which Brian meant, by his utterance, to convey. This example follows the lines laid down by Kripke exactly.

---

[18] Wettstein has emphasized that this is what purports to be the crucial contrast in his article, "Demonstrative Reference and Definite Descriptions," *Philosophical Studies*, XL (1981): 241–57

Moreover, it is my opinion that there are "referential" cases that also give rise to the same pragmatic effects. Suppose that a pair of people, *S* and *S**, know a certain man, Jones, reasonably well. Although he is a familiar acquaintance, they happen not to have learned his name and know him and standardly refer to him as "Muriel's husband." The phrase 'Muriel's husband' is, as it were, a description *of* Jones *for* *S* and *S**. (We also assume that Jones *is* Muriel's husband.) Finally, we add that *S* and *S** also know that Muriel's husband is notoriously lax as a gardener and almost never does any lawn work. The following exchange now takes place:

S: (pointing into the distance) There is Muriel's husband over there.
*S**: What is he doing?
S: You'll never believe this Muriel's husband is raking leaves.

This exchange can be situated in variant contexts in such a way that the cases thus generated will be either simple or complex. It is simple if the man in the distance is Jones, i.e., Muriel's husband. But, if the man in the distance is actually Smith, whom, perhaps because of the bad light, the speakers have mistakenly taken to be Jones, then the case is a complex one. (It will be obvious that this has been constructed in analogy to Kripke's "name" example.)

In either of these cases, on the view I have defended, *S*'s concluding words say literally of Jones—under the correct presupposition that he is Muriel's husband—that he is raking leaves. And, in the complex case, those same words do *not* say literally of Smith that he is raking leaves. When Smith is the man in the distance what *S*'s words express in the context is true if Jones, even in some other locale, is raking leaves. Naturally, in producing his utterance, *S* has, in some sense, said of Smith that he is raking leaves, but that proposition, as before, is simply a part of what *the speaker* meant on the given occasion. The pragmatic apparatus applies here just as it did when proper names were on the stage. It should be agreed on all sides that none of this motivates any further semantical distinctions. The meaning of an attributive description and the meaning of a pronominal description remains unaffected as we pass, in each instance, from simple to complex cases. And this is wholly consistent with what Kripke has showed us concerning the use of proper names.

Entwined in many of the issues considered here are questions about what notion of "singular reference" is taken to be crucial for semantics and about how the favored notion is to be construed. I shall conclude with some brief remarks on this elusive subject. When Kripke introduces the concept of "semantic reference," he stipulates that the semantic referent of a name is its bearer and that the semantic referent of a definite description is the object that uniquely satisfies the descriptor. Taken together, these stipulations imply that it is the semantic function of a description in normal contexts to refer or purport to refer to its unique satisfier. And it is presupposed that 'refers' applies unequivocally to descriptions and proper names. If, however, attributive definite descriptions are quantifier phrases, as I have suggested, then these assumptions may well be wrong. It seems clear in any case that Donnellan intended to reject them in passages like this: "If there is anything which might be identified as reference here [in attributive uses], it is reference in a very weak sense—namely, reference to *whatever* is the one and only one ϕ."[19]

It should be admitted that the import of this passage is cloudy, since the supposed "weak" and "strong" sense of 'refers' is not really delineated. Kripke, in his article, has some excellent remarks about the dangers of postulating an ambiguity when the only grounds for this are that the difference in senses is needed to save some conceptual claim (*op. cit.*, p. 19). One can easily wonder whether Donnellan is not guilty of such a move in the claim cited above, but, although the suspicion is reasonable, I think that there is a simple defense of Donnellan's intuitions.

The verb 'refers' in English is plainly ambiguous in a relevant way. For example, we say

> The phrase 'a stockbroker' *refers* to any person who professionally executes orders to buy and sell securities.

However much philosophers of language may deplore this usage, it is a common one, and its meaning is clear enough: it means "correctly applies to" or "is true of." Adapting some terminology from Russell, let us call "referring" in this sense "denoting." By contrast, if we have a statement with a pure demonstrative, e.g., This is ψ, then we say that the relevant token of 'this' *refers to* such-and-such an object. In making such a report, we do not mean that the word 'this' correctly applies to or is true of the object referred to. Indeed, it is rather difficult to say what exactly we do mean. Very roughly, the token of the demonstrative, construed in relation to other features of the utterance, serves to signal that the speaker's words are meant to say, literally and expressly, *of* some contextually indicated object that *it* is ψ. The pure demonstrative thus systematically contributes to the expression of what is meant to be a singular proposition that contains the referent of the demonstrative as a constituent. Appropriating Russell's terminology once more, let us call reference in this sense "designating."[20]

---

[19] "Reference and Definite Descriptions," p 302. He seems to suggest that it is a mark of an attributive use of $\ulcorner$the $\phi\urcorner$ that the relevant utterance can be paraphrased as $\ulcorner$. . . the ϕ, whoever he (she, it) is, . . $\urcorner$. Certainly many subsequent writers have supposed that this possibility of appositive interjection provides a criterion of attributive use. The test is a bad one, however, and it has led to confusion. I can appropriately assert, while demonstrating a salient individual in front of me, "That man, whoever he is, is a loud buffoon," if I fail to possess whatever interest-relative identificatory information for which the situation and my audience call. Surely my appositive *disclaimer does not make my* use of the demonstrative '*nonreferential*' On the other hand, we may notice that 'whoever', 'whichever', etc., sometimes function as quantifiers, e.g., in Frege's curious example, "Whoever touches pitch defiles himself." It would be a better test of attributive use to ask whether an utterance of $\ulcorner$The ϕ is $\psi\urcorner$ can be paraphrased as $\ulcorner$Whatever is the ϕ is $\psi\urcorner$, where 'whatever' is a quantifier or part of a quantifier phrase. The present quote from Donnellan should be read in this manner.

With this distinction loosely drawn, let me restate what I take to be Donnellan's idea in the quotation set out above. A definite description used attributively refers or purports to refer to its unique satisfier only in the "denoting" sense of 'refers'. The description purports to apply correctly or to be true of at most one object. But, such a description does not designate or purport to designate either the unique satisfier or any other object. The relevant utterance expresses a general proposition, and, therefore, designation of an object by the description is not semantically in play. In the context of his article as a whole, Donnellan is urging that we easily and naturally confuse this point because of the double use of 'refers' and because the very same expression type can also function as a pronominal description used deictically. In the latter case, the description *does* purport to designate an item that satisfies the associated descriptor. I take this interpretation of Donnellan to be confirmed by the substantial emphasis he places on the concept of "saying such-and-such *of* some object" in his discussions of referential use. Thus, in a passage just prior to the one quoted above, he states:

> We saw that when a definite description is used referentially, a speaker can be reported as having said something *of* something. And in reporting what it was of which he said something we are not restricted to the description he used, or synonyms of it; we may ourselves refer to it using any descriptions, names, and so forth, that will do the job (*op. cit.*, p. 303).

One might wish that the passage said more plainly that it is not just the speaker, but also his utterance, which says something *of* something, but here and elsewhere in his article it appears that "referential use" is to be understood in terms of reference qua designation. Reference qua denotation is relevant principally as a source of confusion in discussions of the topic.

---

[20] I have chosen "denoting" and "designating" here because they are suggestive in relation to Russell's use of the terms. But what I say is not at all meant as an explication of *his* concepts.

Definite descriptions occupy a curiously schizophrenic position within the realm of "singular" or "referring" terms. In attributive use, they may refer to an object, but they refer only in the sense that other predicables do: they are true of the object. Descriptions that are narrowly referential (pronominal descriptions used deictically) also refer, but this is the "reference" of names and demonstratives. If $\ulcorner\text{the } \phi\urcorner$ saturates $\ulcorner\text{is } \psi\urcorner$, the whole may be used to say of a contextually determined item o that it is ψ. It is little wonder that the "referential" schizophrenia now diagnosed, has produced a parallel schizophrenia in the history of the topic.

Gottlob Frege subsumed definite descriptions within his broader category of "proper names" and took their referents, if any, to be the unique object described. In fact, he tends implicitly to treat attributively used descriptions as paradigms of the class of proper names. This creates pressures, at least in the exposition of his doctrines, to assimilate the relation of a name to its bearer to the relation of a description to its unique satisfier. The consequent tensions in his system have become well-known. Thus, in a certain system-relative sense, definite descriptions *are* "referring expressions" for Frege. Russell, we noted, divides "reference" into denotation and what he calls "direct designation." The latter is a relation which (probably) demonstratives bear to the objects demonstrated and which "logically proper names" would have to their bearers if any names were ever logically proper. Descriptions, though they may be said to denote, never directly designate anything. So, in this special Russellian sense, descriptions are not "referring expressions" for him. Peter Strawson insisted again that they were. But, on the whole, he tends to veer between the basic Fregean model and a model that stresses deictic use without an accompanying gloss on the attributive features that Russell, in effect, had stressed. It was Donnellan who first saw that there was something radically wrong with the whole debate. He had the insight that each theory was both highlighting one use over the other while, at the same time, muddling properties of the different uses together. Although Donnellan did not bring out the full generality of the linguistic phenomena he observed and was self-admittedly unsure about the semantical status of the two uses, his insight is central for the theory of reference and should not, as it might, be lost.

GEORGE M WILSON

Johns Hopkins University

NEW BOOKS

ACHINSTEIN, PETER: *Particles and Waves: Historical Essays in the Philosophy of Science.* New York: Oxford, 1991. 337 p. Cloth $49.95, paper $24.95.

ACKERMANN, ROBERT JOHN: *Nietzsche: A Frenzied Look.* Amherst: Massachusetts UP, 1990. xi, 205 p. Cloth $24.95.

ADDELSON, KATHRYN PYNE: *Impure Thoughts: Essays on Philosophy, Feminism, and Ethics.* Philadelphia: Temple, 1991. vi, 257 p. Cloth $27.95.

ALLISON, HENRY E.: *Kant's Theory of Freedom.* New York: Cambridge, 1990. xii, 304 p. Cloth $49.50, paper $15.95.

ATWELL, JOHN E.: *Schopenhauer: The Human Character.* Philadelphia: Temple, 1990. xi, 259 p. Cloth $39.95.

BAIER, ANNETTE C.: *A Progress of Sentiments: Reflections on Hume's* Treatise. Cambridge: Harvard, 1991. xi, 333 p. Cloth $39.95.

BARTKY, SANDRA LEE. *Feminism and Domination: Studies in the Phenomenology of Oppression.* New York: Routledge, 1990. xiii, 141 p. Paper $12.95.

BARWISE, JON, and JOHN ETCHEMENDY: *The Language of First-Order Logic: Including the Program* Tarski's World. Stanford: CSLI, 1990. xiii, 259 p. Paper $27.50 (with software).

BERLEANT, ARNOLD: *Art and Engagement.* Philadelphia: Temple, 1991. xvii, 259 p. Cloth $49.95.

BERNADETE, SETH: *The Rhetoric of Morality and Philosophy: Plato's* Gorgias *and* Phaedrus. Chicago: University Press, 1991. ix, 205 p. Cloth $29.95.

BRAMS, STEVEN J.: *Negotiation Games: Applying Game Theory to Bargaining and Arbitration.* New York: Routledge, 1990. xviii, 297 p. Cloth $49.50, paper $15.95.

BROADIE, SARAH: *Ethics with Aristotle.* New York: Oxford, 1991. xiii, 462 p. Cloth $55.00.

BURNYEAT, MYLES: *The* Theaetetus *of Plato.* With a translation by M. J. LEVETT. Indianapolis: Hackett, 1990. xiv, 351 p. Cloth $32.50, paper $10.95.

CARR, DAVID: *Educating the Virtues: An Essay on the Philosophical Psychology of Moral Development and Education.* New York: Routledge, 1991. x, 294 p. Cloth $55.00.

CARRIER, DAVID: *Principles of Art History Writing.* University Park: Pennsylvania State UP, 1991. xiii, 249 p. Cloth $29.95.

CAVELL, STANLEY: *Conditions Handsome and Unhandsome: The Constitution of Emersonian Perfectionism.* Chicago: University Press, 1990 xxxix, 151 p. Paper $8.95.

COLLINS, H. M.: *Artificial Experts: Social Knowledge and Intelligent Machines.* Cambridge: MIT, 1990. xiii, 266 p. Cloth $19.95.

COOPER, DAVID E.: *Existentialism: A Reconstruction.* Cambridge: Blackwell, 1990. ix, 201 p. Cloth $52.95, paper $17.95.

COTHEY, A. L.: *The Nature of Art.* New York: Routledge, 1990. xi, 201 p. Cloth $74.50.

CRAIG, EDWARD: *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. New York: Oxford, 1990. x, 171 p. Cloth $49.95.

CRIMMINS, JAMES E.: *Secular Utilitarianism: Social Science and the Critique of Religion in the Thought of Jeremy Bentham*. New York: Oxford, 1990. xi, 348 p. Cloth $68.00.

DAVIES, STEPHEN: *Definitions of Art*. Ithaca: Cornell, 1991. x, 243 p. Cloth $31.50, paper $12.95.

DOUBLE, RICHARD: *The Non-Reality of Free Will*. New York: Oxford, 1990. xi, 247 p. Cloth $32.50.

DREYFUS, HUBERT L.: *Being in the World: A Commentary on Heidegger's* Being and Time, Division I. Cambridge: MIT, 1991. xiv, 370 p. Cloth $30.00, paper $15.95.

DRYZEK, JOHN S.: *Discursive Democracy: Politics, Policy, and Political Science*. New York: Cambridge, 1990. x, 254 p. Cloth $39.50.

DUMMETT, MICHAEL: *The Logical Basis of Metaphysics*. Cambridge: Harvard, 1991. xi, 355 p. Cloth $34.95.

EAGLETON, TERRY: *Ideology: An Introduction*. New York: Verso, 1991. xv, 242 p. Cloth $59.95, paper $17.95.

ECO, UMBERTO: *The Limits of Interpretation*. Bloomington: Indiana UP, 1990. 295 p. Cloth $27.50.

FENVES, PETER D.: *A Peculiar Fate: Metaphysics and World-History in Kant*. Ithaca: Cornell, 1991. xii, 306 p. Cloth $36.50.

FEREJOHN, MICHAEL: *The Origins of Aristotelian Science*. New Haven: Yale, 1991. ix, 174 p. Cloth $24.50.

FETZER, JAMES H.: *Philosophy and Cognitive Science*. New York: Paragon, 1991. xvii, 170 p. Paper $13.95.

FEYERABEND, PAUL: *Three Dialogues on Knowledge*. Cambridge: Blackwell, 1991. 167 p. Cloth $42.95, paper $17.95.

FLANAGAN, OWEN: *Varieties of Moral Personality: Ethics and Psychological Realism*. Cambridge: Harvard, 1991. xiv, 393 p. Cloth $34.95.

FRONGIA, GUIDO, and BRIAN MCGUINNESS: *Wittgenstein: A Bibliographical Guide*. Cambridge: Blackwell, 1990. x, 438 p. Cloth $65.00.

GELVEN, MICHAEL: *Truth and Existence: A Philosophical Inquiry*. University Park: Pennsylvania State UP, 1990. ix, 204 p. Cloth $25.00.

GORSHOLZ, EMILY R.: *Cartesian Method and the Problem of Reduction*. New York: Oxford, 1991. ix, 161 p. Cloth $42.50.

GOSLING, JUSTIN: *Weakness of the Will*. New York: Routledge, 1990. ix, 221 p. Cloth $39.95.

HACKER, P. M. S.: *Wittgenstein: Meaning and Mind: Volume 3 of an Analytic Commentary on the* Philosophical Investigations. Cambridge: Blackwell, 1990. xxi, 575 p. Cloth $95.00.

HACKING, IAN: *The Taming of Chance*. New York: Cambridge, 1990. xiii, 264 p.,Cloth $44.50, paper $14.95.

HATFIELD, GARY: *The Natural and the Normative: Theories of Spatial Perception from Kant to Helmholtz*. Cambridge: MIT, 1990. xii, 366 p. Cloth $35.00.

HELLER, AGNES: *A Philosophy of Morals*. Cambridge: Blackwell, 1990. xiv, 245 p. Cloth $39.95. And *Can Modernity Survive?* Berkeley: California UP, 1990. vi, 177 p. Cloth $25.00.

HELLER, MARK: *The Ontology of Physical Objects: Four-Dimensional Hunks of Matter*. New York: Cambridge, 1990. xiv, 162 p. Cloth $34.50.

HONDERICH, TED: *Conservatism*. Boulder: Westview, 1990. ix, 255 p. Cloth $45.00.

HOOKWAY, CHRISTOPHER: *Scepticism*. New York: Routledge, 1990. xii, 251 p. Cloth $49.95.

HUNT, LESTER H.: *Nietzsche and the Origin of Virtue*. New York: Routledge, 1991. xxiii, 200 p. Cloth $49.95.

HYLTON, PETER: *Russell, Idealism, and the Emergence of Analytic Philosophy*. New York: Oxford, 1990. xvii, 420 p. Cloth $59.00.

IHDE, DON: *Instrumental Realism: The Interface between Philosophy of Science and Philosophy of Technology*. Bloomington: Indiana UP, 1991. xiv, 159 p. Cloth $29.95, paper $12.50.

INGRAM, DAVID: *Critical Theory and Philosophy*. New York: Paragon, 1990. xxx, 240 p. Paper $12.95.

JACKENDOFF, RAY: *Semantic Structures*. Cambridge: MIT, 1990. xiv, 322 p. Cloth $34.95.

JAMESON, FREDRIC: *Postmodernism, or, the Cultural Logic of Late Capitalism*. Durham: Duke, 1990. xxii, 438 p. Cloth $34.95.

KATZ, JERROLD: *The Metaphysics of Meaning*. Cambridge: MIT, 1990. xiii, 356 p. Cloth $29.95.

KEMP, ANTHONY: *The Estrangement of the Past: A Study in the Origins of Modern Historical Consciousness*. New York: Oxford, 1991. xi, 228 p. Cloth $32.50.

KITCHER, PATRICIA: *Kant's Transcendental Psychology*. New York: Oxford, 1990. xiii, 296 p. Cloth $35.00.

KLEIN, MARTHA: *Determinism, Blameworthiness, and Deprivation*. New York: Oxford, 1990. vi, 232 p. Cloth $39.95.

KRAYNAK, ROBERT P.: *History and Modernity in the Thought of Thomas Hobbes*. Ithaca: Cornell, 1990. vii, 224 p. Cloth $28.50.

KUKATHAS, CHANDRAN, and PHILIP PETTIT: *Rawls*: A Theory of Justice *and Its Critics*. Stanford: University Press, 1990. xi, 169 p. Cloth $22.50, paper $8.95.

KYBURG, HENRY E.: *Science and Reason*. New York: Oxford, 1990. x, 283 p. Cloth $35.00.

LAUDAN, LARRY: *Science and Relativism: Some Key Controversies in the Philosophy of Science*. Chicago: University Press, 1990. xiii, 180 p. Cloth $32.00, paper $12.95.

LEHRER, KEITH: *Metamind*. New York: Oxford, 1990. viii, 313 p. Cloth $39.95.

LEVINSON, JERROLD: *Music, Art, & Metaphysics: Essays in Philosophical Aesthetics*. Ithaca: Cornell, 1990. xi, 419 p. Cloth $49.95, paper $16.95.

LEWIS, DAVID: *Parts of Classes*. Appendix by JOHN P. BURGESS, A. P. HAZEN, and DAVID LEWIS. Cambridge: Blackwell, 1991. x, 155 p. Cloth $39.95, paper $12.95.

LIPTON, PETER: *Inference to the Best Explanation*. New York: Routledge, 1991. x, 194 p. Cloth $49.95.

LLOYD, G. E. R.. *Demystifying Mentalities*. New York: Cambridge, 1990. viii, 174 p. Cloth $49.50, paper $14.95.

MACINTYRE, ALASDAIR: *Three Rival Versions of Moral Enquiry: Encyclopedia, Genealogy, and Tradition*. Notre Dame: University Press, 1990. x, 241 p. Cloth $24.95.

MCGINN, COLIN: *The Problem of Consciousness: Essays towards a Resolution*. Cambridge: Blackwell, 1991. viii, 216 p. Cloth $47.95.

MCINERNEY, PETER K.: *Time and Experience*. Philadelphia: Temple, 1991. vii, 283 p. Cloth $39.95.

MARTIN, MICHAEL: *The Case against Christianity*. Philadelphia: Temple, 1991. xi, 273 p. Cloth $44.95.

MEINWALD, CONSTANCE C.· *Plato's* Parmenides. New York: Oxford, 1991. vii, 192 p. Cloth $32.50.

MICHALSON, GORDON E.: *Fallen Freedom: Kant on Radical Evil and Moral Regeneration*. New York: Cambridge, 1990. xii, 172 p. Cloth $34.50.

MISAK, C. J.: *Truth and the End of Inquiry: A Peircean Account of Truth*. New York: Oxford, 1991. xiii, 182 p. Cloth $49.95.

MORTON, ADAM: *Disasters and Dilemmas: Strategies for Real-life Decision Making*. Cambridge: Blackwell, 1991. xii, 209 p. Cloth $49.95.

NEUHOUSER, FREDERICK: *Fichte's Theory of Subjectivity*. New York: Cambridge, 1990. x, 180 p. Cloth $39.50, paper $12.95.

NUSSBAUM, MARTHA C.: *Love's Knowledge: Essays on Philosophy and Literature*. New York: Oxford, 1990. xix, 403 p. Cloth $42.50.

PARSONS, TERENCE: *Events in the Semantics of English: A Study in Subatomic Semantics*. Cambridge: MIT, 1990. xviii, 334 p. Cloth $35.00.

PEARS, DAVID: *Hume's System: An Examination of the First Book of His Treatise*. New York: Oxford, 1990 ix, 204 p. Cloth $49.95.

POWELL, C THOMAS: *Kant's Theory of Self-Consciousness*. New York: Oxford, 1990. xi, 268 p. Cloth $56.00.

PREVOST, ROBERT: *Probability and Theistic Explanation*. New York: Oxford, 1990. v, 195 p. Cloth $59.00.

PRIOR, WILLIAM J.: *Virtue and Knowledge: An Introduction to Ancient Greek Ethics*. New York: Routledge, 1991. xi, 240 p. Cloth $59.95, paper $15.95.

ROSENFELD, MICHEL. *Affirmative Action and Justice: A Philosophical and Constitutional Inquiry*. New Haven: Yale, 1991. ix, 373 p. Cloth $30.00.

RUBEN, DAVID-HILLEL: *Explaining Explanation*. New York: Routledge, 1990. xi, 265 p. Cloth $39.95.

SCHAEFER, DAVID LEWIS: *The Political Philosophy of Montaigne*. Ithaca: Cornell, 1990. xv, 407 p. Cloth $41.50, paper $13.95.

SCHEFFLER, ISRAEL: *In Praise of the Cognitive Emotions and Other Essays in the Philosophy of Education*. New York: Routledge, 1991. xii, 174 p. Cloth $39.95, paper $13.95.

STATEN, HENRY: *Nietzsche's Voice*. Ithaca: Cornell, 1990. xii, 223 p. Cloth $29.95

STERELNY, KIM: *The Representational Theory of Mind: An Introduction*. Cambridge: Blackwell, 1990. xiii, 252 p. Cloth $44.95, paper $16.95.

STOVE, DAVID: *The Plato Cult and Other Philosophical Follies*. Cambridge: Blackwell, 1991. xiii, 209 p. Cloth $26.95.

THOMSON, JUDITH JARVIS: *The Realm of Rights*. Cambridge: Harvard, 1990. viii, 383 p. Cloth $35.00.

TITO, JOHANNA MARIA: *Logic in the Husserlian Context*. Evanston: Northwestern, 1990. 291 p. Cloth $36.95.

TORRETTI, ROBERTO: *Creative Understanding: Philosophical Reflections on Physics*. Chicago: University Press, 1990. xvi, 369 p. Cloth $79.95, paper $32.50.

TURBAYNE, COLIN MURRAY: *Metaphors of the Mind: The Creative Mind and Its Origins*. Columbia: South Carolina UP, 1991. xv, 133 p. Cloth $24.95.

VAN INWAGEN, PETER: *Material Beings*. Ithaca: Cornell, 1990. viii, 299 p. Cloth $32.50.

WALLER, BRUCE N.: *Freedom without Responsibility*. Philadelphia: Temple, 1990. x, 236 p. Cloth $39.95.

WARDY, ROBERT: *The Chain of Change: A Study of Aristotle's* Physics VII. New York: Cambridge, 1990. x, 345 p. Cloth $59.50.

WEST, CORNEL: *The Ethical Dimensions of Marxist Thought*. New York: Monthly Review, 1991. xxxiv, 183 p. Cloth $36.00, paper $18.00.

WHITE, ALAN: *Within Nietzsche's Labyrinth*. New York: Routledge, 1990. xv, 188 p. Cloth $39.95, paper $13.95.

WILLIAMSON, TIMOTHY: *Identity and Discrimination*. Cambridge: Blackwell, 1990. ix, 173 p. Cloth $45.95.

WOLF, SUSAN: *Freedom within Reason*. New York: Oxford, 1990. xii, 162 p. Cloth $24.95.

WOLFF, JONATHAN: *Robert Nozick: Property, Justice, and the Minimal State*. Stanford: University Press, 1991. xiii, 168 p. Cloth $32.50, paper $9.95.

WOLIN, RICHARD: *The Politics of Being: The Political Thought of Martin Heidegger*. New York: Columbia, 1990. xvii, 221 p. Cloth $29.50.

WOOD, ALLEN W.: *Hegel's Ethical Thought*. New York: Cambridge, 1990. xxi, 293 p. Cloth $49.50.

WYSCHOGROD, EDITH: *Saints and Postmodernism: Revisioning Moral Philosophy*. Chicago: University Press, 1990. xxvii, 298 p. Cloth $40.00, paper $15.95.

YOUNG, IRIS MARION: *Justice and the Politics of Difference*. Princeton: University Press, 1990. viii, 286 p. Cloth $45.00, paper $12.95. And *Throwing Like a Girl and Other Essays in Feminist Philosophy and Social Theory*. Bloomington: Indiana UP, 1990. vii, 213 p. Cloth $29.95, paper $14.50.

BLACKWELL

A COMPANION TO ETHICS
Edited by PETER SINGER
Blackwell Companions to Philosophy Series
April 560 pages
0-631-16211-9 cloth $79.95

THREE DIALOGUES ON KNOWLEDGE
PAUL FEYERABEND
April 204 pages
0-631-17918-6 paper $17.95
0-631-17917-8 cloth $42.95

THE PLATO CULT
And Other Philosophical Follies
DAVID STOVE
March 196 pages
0-631-17709-4 cloth $26.95

MODERNISM AS A PHILOSOPHICAL PROBLEM
On the Dissatisfactions of European High Culture
ROBERT B. PIPPIN
June 224 pages
0-631-17657-8 paper $21.95
0-631-17258-0 cloth $42.95

A PORTRAIT OF WITTGENSTEIN AS A YOUNG MAN
From the Diary of David Hume Pinsent 1912-1914
Edited by
GEORG HENRIK VON WRIGHT
January 120 pages
0-631-17511-3 cloth $29.95

THE BLACKWELL DICTIONARY OF COGNITIVE PSYCHOLOGY
Edited by MICHAEL W. EYSENCK
January 408 pages
0-631-15682-8 cloth $69.95

LOGICAL FORMS
An Introduction to Philosophical Logic
MARK SAINSBURY
April 240 pages
0-631-17778-7 paper $21.95
0-631-17779-9 cloth $64.95

AN INVITATION TO COGNITIVE SCIENCE
JUSTIN LEIBER
June 192 pages
0-631-17005-7 paper $17.95
0-631-17004-9 cloth $49.95

DRETSKE AND HIS CRITICS
Edited by BRIAN McLAUGHLIN
May 438 pages
1-55786-198-6 cloth $47.95

JOHN SEARLE AND HIS CRITICS
Edited by ERNEST LEPORE & ROBERT VAN GULICK
February 420 pages
0-631-15636-4 cloth $44.95

MEANING IN MIND
Fodor and His Critics
BARRY LOEWER & GEORGES REY
June 300 pages
0-631-17103-7 cloth $44.95

BLACKWELL
Publishers

Three Cambridge Center, Cambridge, MA 02142
Call Toll-Free (800) 445-6638

The Idea of Wilderness
From Prehistory to the Age of Ecology

Max Oelschlaeger

In a wide-ranging work of intellectual history that is certain to become a classic, Max Oelschlaeger presents a stimulating new examination of humankind's relationship with the natural world through the ages.

"An absolutely crucial and timely benchmark in ecophilosophic thought and history."—George Sessions $29.95

The Life and Thought of Friedrich Engels
A Reinterpretation

J.D. Hunley

In a major reinterpretation of Engels's life and ideas—focusing in particular on his relationship with Karl Marx—J.D. Hunley shows that Engels did not disagree with Marx about important issues, nor did he distort Marx's views after the latter's death.

"A well-informed and well written biographical and interpretive account of Engels' life and views." —Allen Wood $22.50

The Self in Social Theory
A Psychoanalytic Account of Its Construction in Plato, Hobbes, Locke, Rawls, and Rousseau

C. Fred Alford

Alford sets forth a psychoanalytic account of the self and applies it to texts by Plato, Hobbes, Locke, Rawls, and Rousseau in order to draw out their implicit, often inchoate, assumptions about the self.

"The book is a fine piece of scholarship, elegantly and engagingly written."—Fred Dallmayr $30.00

Difference and Subjectivity
Dialogue and Personal Identity

Francis Jacques
translated by Andrew Rothwell

This book—the first by this important French philosopher to be translated into English—investigates the question of human subjectivity.

"This work, important and rich, intends nothing less than to present a new conception of the speaking and thinking subject."—Yves Michaud, *Libération* (on the French edition) *Winner of the Broquette-Gonin prize of the Academie Francaise* $40.00

Benjamin Constant and the Post-Revolutionary Mind

Biancamaria Fontana

A concise overview of the life and writings of the Swiss political theorist Benjamin Constant, a key figure in the attempt to come to terms with the political world that followed the French Revolution. Fontana shows the unity of Constant's vision and the relevance between the issues he discussed and those that confront modern democratic states today. $25.00

Yale

The Origins of Aristotelian Science

Michael Ferejohn

In this book Michael Ferejohn offers the first comprehensive interpretation of Aristotle's theory of demonstrative knowledge, as that theory is presented throughout most of the *Posterior Analytics* and parts of the *Prior Analytics*.

"Highly original and clearly written, this book provides **a well-defended and novel interpretation** of one of **the most important works** in the history of philosophy."—James Lennox $24.50

Philosophical Hermeneutics and Literary Theory

Joel Weinsheimer

In this lucid and elegantly written book, Joel Weinsheimer discusses how the perceptions of Hans-Georg Gadamer alter our understanding of literary theory and interpretation.

"Weinsheimer ably explores some neglected implications of Gadamer's philosophical hermeneutics. **His well-written book will add new insights** to current debates about the interpretation of texts and the revision of canons."—Paul Hernadi $19.50

Yale University Press
Dept. 646
92A Yale Station
New Haven, CT 06520

Now available in paperback

The Ancients and the Moderns

Rethinking Modernity

Stanley Rosen

The eminent philosopher Stanley Rosen takes a new look at the famous "quarrel" that the moderns have with the ancients.

"**Will make a significant difference in the way we view philosophy over the coming decades.**"—Kenley Dove $12.95

Philosophy in World Perspective

A Comparative Hermeneutic of the Major Theories

David A. Dilworth

"**Provocative and challenging**....It is to be hoped that [this book] will acquaint more Western readers with Chinese philosophy, while spurring Asian thinkers to offer counterproposals about the crucial issues of philosophy in their respective traditions."—Carl Becker, *The Journal of Asian Studies*

"Opens new interpretive possibilities for intra- and inter-textual reflection on a grand scale."—Edith Wyschogrod $14.00

An Interpretation of Religion

Human Responses to the Transcendent

John Hick

"This book strengthens Hick's position as one of the most significant thinkers of the second half of the twentieth century....It is [Hick's] **most comprehensive, bold, and challenging work.**"—Chester Gillis, *Journal of Religion* $18.95

New Perspectives from MIT

Studies in Contemporary German Social Thought

The Persistence of Modernity

Essays on Aesthetics, Ethics, and Postmodernism

Albrecht Wellmer

Albrecht Wellmer defends the general thesis that modernity contains its own critique and that what has been called postmodernism is in fact a further articulation of that critique.

GST 272 pp. $24.95 (July)

Adorno's Aesthetic Theory

The Redemption of Illusion

Lambert Zuidervaart

This is the first book to put *Aesthetic Theory* into context and outline the main ideas and relevant debates, offering readers a valuable guide through this huge, difficult, but revelatory work.

GST 472 pp. $37.50 (July)

Now Available in Paperback

Political Romanticism

Carl Schmitt

translated by Guy Oakes

GST 277 pp. $9.95

To order call toll-free 1-800-356-0343
or (617) 625-8569
MasterCard and Visa accepted

Communicative Action

Essays on Jürgen Habermas's *The Theory of Communicative Action*

edited by Axel Honneth and Hans Joas

translated by Jeremy Gaines and Doris L. Jones

"Habermas's monumental *Theory of Communicative Action*, considered by many the summa of his remarkable career, engendered an intense international debate. Many of the most trenchant contributions have been judiciously assembled by Axel Honneth and Hans Joas for this volume. The result is a book that is not merely about communicative action, but also its very embodiment."—Martin Jay, University of California

GST 381 pp. $39.95

Now Available in Paperback

On Walter Benjamin

Critical Essays and Recollections

edited by Gary Smith

GST 400 pp. $14.95

Nietzsche and Political Thought

Mark Warren

GST 311 pp. $12.50

Ideals and Illusions

On Reconstruction and Deconstruction in Contemporary Critical Theory

Thomas McCarthy

"At a time when debates concerning modernity and postmodernity have become so polemical, Thomas McCarthy displays the virtue of judicious critique.... With this book, McCarthy's own 'voice'–his distinctive contribution to the ongoing project of critical theory–emerges with clarity"— Richard J. Bernstein, New School for Social Research

254 pp. $22.50

Now Available in Paperback

French Modern

Norms and Forms of the Social Environment

Paul Rabinow

454 pp. $14.95 paper

ZONE BOOKS in paperback

Masochism
Coldness and Cruelty
Gilles Deleuze

Venus in Furs
Leopold von Sacher-Masoch

The Accursed Share
Volume 1
Georges Bataille
translated by Robert Hurley

Myth and Tragedy in Ancient Greece
Jean-Pierre Vernant and Pierre Vidal-Naquet

Myth and Society in Ancient Greece
Jean-Pierre Vernant

Matter and Memory
Henri Bergson

Bergsonism
Gilles Deleuze

The MIT Press
55 Hayward Street
Cambridge, MA 02142

NEW PERSPECTIVES FROM MIT/BRADFORD BOOKS

Original in Paperback

THE PHILOSOPHY OF SCIENCE

edited by Richard Boyd, Philip Gasper, and J.D. Trout

The more than forty readings in this anthology chart the rise and decline of logical positivism and the gradual emergence of a new consensus. As an introduction to the philosophy of science, it stands out for its scope, its coverage of both historical and contemporary developments, and its detailed introductions to each area discussed.

A Bradford Book 816 pp., 18 illus. $45.00 cloth, $22.50 paper

Now Available in Paperback

RECONSTRUCTING THE PAST

Parsimony, Evolution, and Inference

Elliott Sober

"Sober is a philosopher who studies biology, and he brings a rare skill to bear upon the philosophical arguments with which biologists (and others) have tried to justify parsimony."— Mark Ridley, *Nature*

A Bradford Book 265 pp. $14.95 paper

MODULARITY IN KNOWLEDGE REPRESENTATION AND NATURAL-LANGUAGE UNDERSTANDING

edited by Jay L. Garfield

A Bradford Book 427 pp. $14.95 paper

MEANING AND MENTAL REPRESENTATION

Robert Cummins

A Bradford Book 180 pp. $9.95 paper

THE SCIENCE OF THE MIND

Second Edition

Owen Flanagan

"Flanagan has done it again. He has the rare skill of taking complex ideas and rendering them in a clear and concise narrative. Any student of the mind sciences should read this book."— Michael S. Gazzaniga, Dartmouth Medical School, Department of Psychiatry

A Bradford Book 440 pp. $27.50 cloth, $14.50 paper

EXPLAINING BEHAVIOR

Reasons in a World of Causes

Fred Dretske

"[An] important book. Dretske is well-informed in cognitive science, evolutionary biology and animal behavior.... His careful, non-technical presentation will be a focus for discussion in the philosophy of mind."— Christopher Hookway, *Times Higher Education Supplement*

A Bradford Book 165 pp. $9.95 paper

MICROCOGNITION

Philosophy, Cognitive Science, and Parallel Distributed Processing

Andy Clark

Explorations in Cognitive Science. *A Bradford Book* 226 pp. $10.95 paper

THE MIT PRESS

Philosophy

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NUMBER 8
AUGUST 1991

page
393 *Holism and Nonseparability* — Richard A. Healey

BOOK REVIEWS
422 Charles Taylor: *Sources of the Self* — J. B. Schneewind
427 William Hasker: *God, Time, and Knowledge* and Edward R. Wierenga: *The Nature of God: An Inquiry into Divine Attributes* — John Martin Fischer
434 Judea Pearl: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* — Henry E. Kyburg, Jr.
437 Peter Gärdenfors: *Knowledge in Flux: Modeling the Dynamics of Epistemic States* — Isaac Levi

445 New Books: Anthologies
446 New Books: Paperbacks
447 New Books: Translations

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1991 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices
Send address changes to *Journal of Philosophy*, 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NO. 8, AUGUST 1991

HOLISM AND NONSEPARABILITY*

Metaphysics follows physics in fact as well as in Aristotle. This is hardly a novel insight as applied to the views of great philosophers of the past, such as Descartes, Leibniz, or Kant. But it holds morals for contemporary thought also. One moral is that much contemporary metaphysics, rather than proceed independently of any particular view of the physical world, in fact rests on uncritical assumptions as to its nature which are hard to reconcile with contemporary physics. A second moral is that radical developments in physics, such as the theories of relativity and quantum mechanics, are themselves a rich source of metaphysical insights, irrespective of the ultimate fate of those theories or of particular interpretations of them.

The claim that quantum phenomena exhibit a characteristic holism or nonseparability provides an important illustration of both morals. Niels Bohr and Erwin Schrödinger put forward such ideas in the early days of quantum theory, but they have been repeated more insistently in recent years by a number of physicists and philosophers.[1] Metaphysicians cannot afford to ignore these ideas. But this

---

* For constructive comments on an earlier version of this paper, I thank Michael Jubien, Paul Teller, Martin Jones, and other members of our Northern California discussion group, as well as the philosophy department at the University of Arizona.

[1] Bohr often remarked upon the holistic character of quantum descriptions. He stressed the fact that such descriptions are significant only in the context of an entire experimental arrangement, which he took to be required for the meaningful attribution of properties to its elements, including the atomic objects on which it may be used to perform measurements. Schrödinger, on the other hand, emphasized the idea of nonseparability. He stressed the fact that individual quantum systems often fail to have separate quantum states of their own, even though the compound quantum system they compose does have such a state. See Bohr, *Atomic Theory and the Description of Nature* (New York: Cambridge, 1934); Schrödinger, "Discussion of Probability Relations between Separated Systems," *Proceedings of the Cambridge Philosophical Society*, XXXI (1935): 555–63.

0022-362X/91/8808/393–421

does not mean that claims of holism or nonseparability should be accepted without subjecting them to detailed critical analysis. It is the purpose of this paper to contribute to such an analysis. Only when we know just what the theses of holism and nonseparability amount to will we be in a position to assess the extent to which quantum phenomena establish these theses and to explore their further metaphysical implications.

This task is made more difficult but at the same time more important by many philosophers' natural suspicions of holism. Such suspicions may even be strengthened by a survey of previous attempts to characterize quantum holism.[2] Clean and interesting holist theses need to be formulated if the concept is to be rehabilitated.

I shall offer formulations of both holism and nonseparability. These turn out to be interestingly different, but closely related notions. Indeed, they are so closely related that it is not surprising that previous analyses have not adequately distinguished them. Whether quantum mechanics establishes either holism or nonseparability depends on how one interprets quantum mechanics. But the metaphysician should not use controversy about the correct interpretation of quantum mechanics as an excuse for not investigating the implications of holism and nonseparability; for quantum mechanics comes about as close to demonstrating holism and nonseparability as it does to establishing indeterminism. Just as in the case of indeterminism, the prospect of holism and nonseparability presents a stimulus, not a threat, to metaphysics.

I

Holism is often characterized as the view that the whole is more than the sum of its parts. This has a paradoxical ring to it, since on one natural understanding it is obviously false: any whole consists precisely of its parts, no more and no less. If the whole seemed to be more than the sum of its parts, then that could only be because additional parts were being overlooked. Nevertheless, it will serve as a suggestive metaphor guiding the search for more acceptable formulations of holism.

---

[2] This is particularly true of Bohr's notoriously obscure discussion of holism. But more recent discussions of holism and nonseparability have been more suggestive than precise. See, among others, David Bohm, *Wholeness and the Implicate Order* (New York Routledge & Kegan Paul, 1980); Bernard d'Espagnat, *In Search of Reality* (New York. Springer, 1983); Don Howard, "Einstein on Locality and Separability," *Studies in History and Philosophy of Science*, XVI (1985): 171–201, and "Holism, Separability and the Metaphysical Implications of the Bell Experiments," in J. T. Cushing and E. McMullin, eds., *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem* (Notre Dame: University Press, 1989); Michael Redhead, *Incompleteness, Nonlocality and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics* (New York: Oxford, 1987). Paul Teller has proposed what he calls "relational holism" as an explication of the kind of holism or nonseparability evidenced in the quantum domain. His analyses may be fruitfully compared with those I offer here: see, in particular, his "Relational Holism and Quantum Mechanics," *British Journal for the Philosophy of Science*, XXXVII (1986): 71–81, and "Relativity, Relational Holism, and the Bell Inequalities," in J. T. Cushing and E. McMullin, eds. The present paper owes much to the work of these and other recent authors; I have tried to straighten, smooth, and signpost the trail they have blazed.

As a first step, consider the nature of the composition relation. If one were to take this to be fully captured by the axioms of mereology, then to an *arbitrary* subset of objects S from a given set there would correspond a whole $M(S)$—the mereological sum of S. If none of the objects in S overlaps another, the objects in S may be said to compose $M(S)$. Note that, with composition understood in terms of mereology, the objects in S compose the whole $M(S)$ irrespective of their properties or relations. On one natural understanding, holism would then become trivially false; for $M(S)$ is precisely the (mereological) sum of its parts.

We can try to rescue the thesis of holism from the fate of trivial falsity by noting that for something to be a typical whole of a certain kind it is not enough for the mereological sum of its parts to exist. For there to be a model airplane is not just for the mereological sum of its parts to exist at a time, even if these do then compose it. If the builder had left the kit unopened the previous day instead of constructing the model, the mereological sum of its parts would have existed at that time without thereby composing this (or indeed any) airplane. Other examples show that the properties of the individual parts are just as important as their spatial relations to their composing a whole of a certain kind. A particular object would not have been the nuclear bomb that it is if the atoms of uranium 235 which it contains had had the same half-life (4.5 billion years) as atoms of uranium 238.

Clearly many composite objects are in some sense more than just mereological sums of their parts. But in what sense? Call any part of an object which is neither identical to that object, nor a part of every object, a *proper* part. An object is *composite* just in case it has proper parts. Two proper parts *overlap* if one is a part of the other or if there is an object that is a proper part of both of them. A set of proper parts of an object no two of which overlap is a *nonoverlapping set*. A nonoverlapping set P of proper parts of an object o such that every part of o overlaps some element of P *composes* o: the objects in P are then *component parts* of o (relative to the partition P, though the qualification may often be dropped without confusion). We may call a composite object o a *genuine whole with component parts* P if P composes o, but the existence of each element of P is not a sufficient condition for the existence of o.

There is then a trivial sense in which every genuine whole is more than the sum of its parts; for any genuine whole is distinct from the mereological sum of its components. All that is required for the existence of the mereological sum is the existence of the elements of *P*. But this does not suffice for the existence of the genuine whole. For the elements of *P* compose *o* only if each has the required properties, and these elements are appropriately arranged. This is true despite the fact that *o* is composed solely of elements of *P*.

Of course, one may still question whether there are any genuine wholes, as I have defined them. If, for example, one believes that an ordinary object like a model airplane, a table, or a nuclear bomb is identical to a certain mereological sum, then one will not accept that such an object is a genuine whole.[5] If there are no ordinary genuine wholes, one may begin to doubt whether there are any extraordinary ones either. But if in fact there are no genuine wholes, the fact that any genuine whole is more than the sum of its parts becomes even more trivially true!

How can we arrive at a formulation of holism that is neither trivially false nor trivially true? I think the key is to focus on respects in which the features of a whole may be independent of those of its proper parts.

Wholes frequently exhibit features that seem qualitatively distinct from those of their component parts (relative to some partition). The water in my kettle is wet and boils at 100 degrees Celsius, even though none of its component molecules is wet or has a boiling point. And copper is a red metal that conducts electricity, even though none of the protons, neutrons, or electrons that make it up is red or metallic or an electrical conductor. Properties of a whole possessed by none of its component parts (relative to some chosen partition) could be called emergent. One way of understanding holism is as the thesis that a whole of type *W* composed of a set *P* of proper parts is more than the sum of these parts in the sense that any object of type *W* has some property *F* while none of these component parts has property *F*. On this way of understanding it, a compound object could instantiate holism to various degrees, depending firstly on the choice of partition, and secondly on the variety and importance of emergent properties possessed by objects of that type.

---

[5] Jubien, for example, believes this, cf. his (unpublished) manuscript "Ordinary Naming and Necessity." He argues there that, even though an object such as a particular table is identical to a mereological sum, that mereological sum has certain properties, including the property of "being this table," which it might have lacked. On this view, moreover, the object with the property of "being this table" might have been a different mereological sum. Jubien takes this last consequence of his view to be the correct rendering of the intuition that this table might have had different parts from those which it in fact has. He is thus able to acknowledge a sense in which a certain *table* could have had different parts, even though the mereological sum that is that table could not. In general, on this view, an object's being a particular whole of a certain kind is not just a matter of its being a certain mereological sum, even though it *is* a mereological sum (and consequently *not* a genuine whole in my sense): for a different mereological sum could have uniquely instantiated the property of being the particular whole of that kind.

But holism in this sense is surely too familiar a feature of the world to occasion much surprise or comment. The attempt to recover all the sensible properties of physical objects starting from a severely restricted set of properties of their constituents has been a familiar theme in metaphysics since the pre-Socratics and in physical science at least since the time of Galileo. In fact, it would be more surprising if such attempts could be shown to be subject to serious obstacles. And this presents a clue to more interesting holist theses. Rather than assert that a whole has features that none of its components has, holism should be understood to claim that the whole has features that cannot be reduced to features of its component parts. But now we need to get clearer on what we mean by 'features' and 'reduction'.

There are at least two different approaches we can take at this point. The first approach leads to formulations of holism best suited for a discussion of explanatory or methodological issues, while the second leads to formulations that are of more direct relevance to metaphysics. Since my interest here is metaphysical, I shall be pursuing the second approach in what follows. But it will help to make this approach clearer if I begin by saying a few words about the kind of holism with which it is to be contrasted. For explanatory holism is perhaps a more familiar thesis, and while it is important to make clear that this is not the kind of holism with which I shall be primarily concerned, the truth of metaphysical holism would have interesting implications for explanatory holism. Epistemology can be divorced from metaphysics no more here than elsewhere.

II

Holism claims that a whole has features that cannot be reduced to features of its parts. If the features of an object include its behavior, and reducing the features of a whole to those of its parts means explaining its behavior in terms of their behavior, then we arrive at a thesis of explanatory holism. Explanatory holism is the view that a satisfactory explanation of the behavior of an object of some type cannot be given by analyzing that object into its component parts and appealing to the laws that hold of these parts. Thus, a biological holist may maintain that it is impossible fully to explain the functioning of an entire organism purely in terms of the physical and chemical laws obeyed by the molecules of which its cells are composed. And a social holist may maintain that an adequate explanation of the behavior of a society will require more than knowledge of the principles governing the actions of the society's individual members.[4]

Explanatory holism thus stands opposed to explanatory reductionism—the view that one can achieve a completely adequate understanding of the behavior of the whole by analyzing it into its component parts and discovering the laws according to which these operate. Such a reductive methodology has proven enormously successful in physical science since the time of Galileo, so that today it is common to assume that elementary particle physics lies at the foundation of all physical science, and that, if only physicists knew the laws governing the fundamental constituents of matter, they would truly have "a theory of everything." Nevertheless, there continue to be influential dissenting voices, even among physicists. Thus in his recent book, *The Problems of Physics,* A.J. Leggett[5] considers the physics of condensed matter. He writes

> . . . there are also a number of cases, particularly in the traditional areas of the physics of gases and crystalline solids, in which a model which treats the behavior of the whole as essentially just the sum of that of its parts (atoms or electrons) has been quite successful. . . . But these cases, despite the fact that they totally dominate the presentation of the subject in most elementary textbooks, are actually the exception rather than the rule. In virtually all the frontier areas of modern condensed-matter physics, the relationship between our understanding of the behavior of matter at the microscopic level of single atoms and electrons, and at the macroscopic level of (say) liquids and solids, is actually a good deal more complicated than this. . . . I would claim that the most important advances in this area come about by the emergence of qualitatively new concepts at the intermediate or macroscopic levels—concepts which, one hopes, will be compatible with one's information about the microscopic constituents, but which are in no sense logically dependent on it (*ibid.,* p. 113).

Even within physics, then, there continues to be debate about the respective merits of explanatory reductionism and explanatory holism. If metaphysical holism is true, then there is at least one domain in which reductive explanations will be unavailable. This may prompt a reassessment of research strategy also in other domains. If so, metaphysics may come to influence methodology.

---

[4] Of course, other varieties of biological and social holism may prove more worthy of debate.

[5] New York: Oxford, 1987.

But just what are the theses of explanatory reductionism and explanatory holism? A somewhat more precise formulation requires preliminary discussion of ultimate and basic parts. An ultimate part of an object is a part that itself has no proper parts. Whether or not an object has ultimate parts is a question for physics, and not just for philosophy. Atoms are no longer considered ultimate constituents of matter by contemporary physics. Although electrons and quarks are better candidates for that title, the discovery of particles at successively more elementary levels below the level of the atom has caused some physicists to question whether matter *has* ultimate constituents, and even if it does, whether we shall ever be in a position to know what they are. But whether or not matter has ultimate parts, we may still wish to speak of the basic parts of an object. For example, even though atoms are not the ultimate constituents of matter, there is still a sense in which they may be claimed to be the basic components of ordinary matter. But in what sense? We may try to capture that sense as follows:

A set B of objects is *basic* relative to a domain D of objects and processes of type P just in case

(1) Every element of D is composed solely of objects in B,
(2) the objects in B do not overlap,
(3) while processes of type P may create or destroy elements of D, or may convert one element of D into another, no process in P can create, destroy, or interconvert elements of B, and
(4) there is no set C satisfying (1)–(3) such that some element of C has some element of B as a proper part.

Intuitively, the basic parts relative to D and P may be thought of as the *building blocks* of elements of D: the largest components that retain their integrity through processes in P capable of building, destroying, and transforming elements of D. A *basic part* of an object o (relative to a domain D and processes P) is just a part of o which is basic relative to D and P.

It is a consequence of these definitions that some basic parts may be ultimate, while others may not be. Ultimate parts of an object are basic relative to any domain of objects that includes it and to any processes that neither create, destroy, nor exchange these ultimate parts. But note that the so-called elementary particles will not be basic parts of all objects relative to all processes; for there are processes that create, destroy, and interconvert elementary particles.

We may now formulate the following theses:

Pure physical explanatory reductionism: The physical behavior of any domain D of physical objects subject only to processes P can be explained in terms of the physical properties and relations of their basic physical parts (relative to D and P).

Pure physical explanatory holism: There is some domain D of physical objects subject only to processes P, not all of whose physical behavior can be explained in terms of the physical properties and relations of their basic physical parts (relative to D and P).

Each of these theses is described as "pure" because each is restricted in its application to physical objects. One might formulate a more ambitious thesis of physical explanatory reductionism according to which the behavior of *every* object can be thus explained. This more ambitious thesis would form a natural supplement to the physicalist view that all objects are, at bottom, physical objects. The pure explanatory theses, on the other hand, are naturally understood to presuppose the following more restricted principle:

Pure physical composition: Every basic part of a physical object (relative to some domain D and processes P) is itself a physical object.

III

The kind of holism I want to focus on now may be called metaphysical rather than explanatory. A thesis of metaphysical holism stands opposed to a claim of supervenience in some domain. The supervenience claim is that all properties of individuals in the domain are determined by the properties of their parts.[6] 'Partism' would be a suggestive, but ugly, name for this claim: in the interests of harmonious usage, I adopt Teller's term 'particularism' instead. But note that a proponent of particularism in this sense need bear no animosity toward universals. It is rather metaphysical holism that the particularist is concerned to deny. The metaphysical holist claims that some individuals in the domain in question have properties that are not determined by the properties of their parts. We need to see how to make these claims more precise in the case of physical holism.

The first point to note is that, if we are to arrive at an interesting formulation of metaphysical holism, we must accept that this thesis is not only concerned with properties, and not concerned with all properties. The properties of a genuine whole will typically depend upon *relations* among its proper parts as well as on properties of the individual parts. But if we are permitted to consider *all* properties and relations among the parts, then these trivially determine the properties of the whole they compose. For one relation among the parts is what we might call the *complete composition relation*—that relation among the parts which holds just in case they compose this very whole with all its properties.

---

[6] While the notion of supervenience is itself not exactly unproblematic (as a large recent philosophical literature attests), an analysis of holism in terms of supervenience may still contribute to philosophical progress. I take the determination involved here to be not logical but physical (in some suitably broad sense).

Let us call the set of properties and relations of the parts upon which the properties and relations of the whole may or may not supervene the ***supervenience basis.*** To avoid trivializing the theses we are trying to formulate, only certain properties and relations can be allowed in the supervenience basis. The intuition as to which these are is simple—the supervenience basis is to include just the qualitative, intrinsic properties and relations of the parts, i.e., the properties and relations that these bear in and of themselves, without regard to any other objects, and irrespective of any further consequences of their bearing these properties for the properties of any wholes they might compose. Unfortunately, this simple intuition resists precise formulation. It is notoriously difficult to say precisely what is meant either by an intrinsic property or relation, or by a purely qualitative property or relation.[7] And the other notions appealed to in expressing the simple intuition are hardly less problematic. But, imprecise as it is, this statement serves already to exclude certain properties and relations from the supervenience basis.

In particular, besides excluding the complete composition relation, it excludes what we might call historical properties and relations, and individual properties and relations. A historical property of an object is a property that object has at a time simply in virtue of what happens to it at some other time. Having been driven 30,000 miles is a historical property of my car. An individual property of an object is a property whose instantiation depends upon the existence of some individual. It is both a historical and an individual property of the Mona Lisa that it was painted by Leonardo da Vinci.

The only properties and relations of the parts which are allowed in the supervenience basis are those which the relevant parts actually manifest at the time in question, regardless of any other individuals, and irrespective of what happens to the parts at other times. Notice that some but not all properties and relations in the supervenience basis may be essential properties of the relevant parts—properties or relations those parts must have if they are to exist at all.

Now it is clear that historical or individual properties of a whole cannot be expected to supervene on a basis of properties and relations of their parts which fails to include their historical or individual properties and relations. Consequently, the twin claims of particularism and metaphysical holism should be restricted so that each applies only to the qualitative, intrinsic properties of the whole. On the other hand, these claims may be broadened so that they apply not merely to a single whole, but also to a set of wholes of some kind. If the claims are so broadened, then the determination of any relations among the elements of the set by the properties and relations of their parts in the supervenience basis is just as much in question as the determination of any of their individual properties.

---

[7] David Lewis explicitly discusses the notion of an intrinsic property, and implicitly also that of an intrinsic relation, in *On the Plurality of Worlds* (New York: Blackwell, 1986). He argues there that appeal to a basic unanalyzed notion of intrinsic property is philosophically legitimate, while also pointing to two alternative attempts to explicate that notion. In his earlier paper, "Extrinsic Properties," in *Philosophical Studies*, XLIV (1983): 197–200, he criticized another, simpler attempted explication of this notion. A property or relation may be said to be qualitative if it does not depend on the existence of any specific individual. But this does not, of course, constitute a satisfactory definition.

Finally, we arrive at the following opposing theses:

> *Pure physical particularism:* Every qualitative, intrinsic physical property and relation of a set of physical objects from any domain *D* subject only to processes of type *P* is supervenient upon the qualitative, intrinsic physical properties and relations of their basic physical parts (relative to *D* and *P*).
>
> *Pure physical holism:* There is some set of physical objects from a domain *D* subject only to processes of type *P*, not all of whose qualitative, intrinsic physical properties and relations are supervenient upon the qualitative, intrinsic physical properties and relations of their basic physical parts (relative to *D* and *P*).

As in the case of the theses of explanatory holism and reductionism considered earlier, each of these theses may be taken to presuppose pure physical composition.[8]

It is easy to give examples that conform to pure physical particularism. The temperature of some particular sample of gas supervenes upon the translational kinetic energies of its constituent molecules. My body's currently being attracted toward the center of the earth with a force of some 160 pounds weight supervenes upon the masses of each part of my body and of the earth, and on the distance between each part of my body and each part of the earth.

But one cannot provide familiar illustrations of pure physical holism, precisely because exceptions to pure physical particularism lie outside the range of our normal physical experience. This is what gives quantum holism metaphysical bite. It is worth considering why it is so difficult to come up with such illustrations. I think the main reason is that physical composition is naturally thought of as resulting from spatial juxtaposition. Even though there are examples of intrinsic relations that are not spatial, such examples seem to depend on intrinsic spatial relations. The size of the electrostatic attraction between two charged particles is an intrinsic relation between the particles which is not a spatial relation. It does, however, depend on the spatial distance between the particles as well as on their individual electric charges. Genuine physical wholes seem built up out of their parts by putting these parts together in the right spatial relations.

---

[8] One might try to explicate the notion of supervenience here in terms of models of a true, descriptively complete, physical theory. At issue is whether there exist two models of such a physical theory of some objects in *D* which agree on the qualitative, intrinsic physical properties and relations of the basic parts of these objects but disagree on some qualitative intrinsic property or relation of the objects themselves.

There are many cases in which this turns out to be true, even when at first sight other causal relations between the parts seem to be playing an independent role in the constitution of the whole. It may seem that all the parts of an amplifier are correctly spatially arranged even though it does not work properly, but closer examination may reveal a short circuit—itself a product of inappropriate spatial relations among the small scale parts of the amplifier. Frequently, intrinsic relations among parts of a genuine whole which do not supervene on intrinsic properties and spatial relations of *those* parts do supervene on intrinsic properties and spatial relations of more elementary parts. We have a strong intuition, based on a great deal of physical experience, that the intrinsic physical properties and relations of composite objects are determined by the intrinsic physical properties and spatial relations of their component parts, even though the parts for which this is true may be microscopic or even ultimate.

It might be objected that the reason it is so difficult to come up with illustrations of physical holism is that there can be none, because this thesis is a priori false. I can think of two arguments for this conclusion, but neither is successful. Here is the first argument. Let w be a particular physical object composed of a set $P(w)$ of basic physical parts, which have the necessary, intrinsic physical properties and relations to compose w. Suppose that w has some physical property J even though the $P(w)$ might have had all the same intrinsic physical properties and relations without its being true that w has J. Now, the $P(w)$ are physically arranged precisely so as to compose w, which has J. Hence the set $P(w)$ of basic physical parts has the necessary intrinsic physical properties and relations to compose w, complete with the property J. Thus J is in fact supervenient upon the intrinsic properties and relations of the $P(w)$, and so pure physical particularism holds after all.

This argument embodies a modal fallacy. It is fallacious to infer from the premise that the $P(w)$ has all the necessary intrinsic physical properties and relations to compose w, which has J, to the conclusion that these properties and relations are incompatible with w's failing to have J. Only if J were an essential property of w would w's existence guarantee that it has J. But not all intrinsic properties are essential. To illustrate pure physical holism, all that is necessary is to find some nonessential intrinsic physical property of a whole which does not supervene on the intrinsic physical properties and relations of its physical parts.

The second argument is not so easy to dispose. Suppose, as before, that w is composed of physical parts $P(w)$, and that w in fact has some nonessential intrinsic property J. There is a relation R among the $P(w)$ which can be defined as follows: R holds among the $P(w)$ if and only if w exists and has J. Moreover, this relation R should be included in the supervenience basis along with any other qualitative, intrinsic physical relations among the $P(w)$. But if R is included in the supervenience basis, then this is after all sufficiently complete to determine not only that w exists but also that w has J. Hence pure physical particularism holds here after all.

The response to this argument is to question the assumption that the relation R should be included in the supervenience basis. Recall that the complete composition relation C was earlier disqualified from appearing in the supervenience basis on the grounds that C is not a qualitative, intrinsic physical relation among the $P(w)$—a relation that holds without regard to any further consequences of its holding for any whole that the $P(w)$ might compose. The relation R should also be excluded from the supervenience basis on these same grounds. It is not clear that the relation R is a physical relation. Even if R is counted as a physical relation, it seems clear that it is not an *intrinsic* physical relation, even though the property J to which it corresponds is an intrinsic physical property of w. R is not a relation that the $P(w)$ have in and of themselves: it is a relation they inherit only insofar as they compose a whole w with property J.[9]

---

[9] In fact, more needs to be said in response to the second argument. For certainly some nonessential, qualitative intrinsic physical properties of some wholes do supervene on the qualitative, intrinsic physical relations as well as properties of their component parts. If J is such an intrinsic property of a whole w, then there exists a qualitative, intrinsic physical relation R' among the $P(w)$ whose holding helps to determine that w has J. Moreover, there seems no a priori justification for excluding cases in which the holding of R' among the $P(w)$ is both necessary and sufficient for w to exist and have J. But in such a case, there is a relation (R') in the supervenience basis which, of necessity, holds among the $P(w)$ if and only if w exists and has J. And this seems to show that the relation R defined earlier cannot, after all, be automatically excluded from the supervenience basis.

Even in such a case, however, the relations R and R' are not identical, and while R' must be included in the supervenience basis, R should not be. For R it is a matter of definition that it holds among the $P(w)$ if and only if w exists and has J, while for R' this is not true by definition, but only as a matter of what might be called natural necessity. In terms of the metaphor of possible worlds, R and R'

Although this response to the second argument seems intuitively powerful, it is not as conclusive as the response to the first argument, because the distinctions on which it rests have not been given precise analyses. Even in the absence of such analyses, the response may be made more persuasive by exhibiting a concrete case of a whole w, a property J, and relation R among the basic parts $P(w)$ of w in which it is intuitively clear that, while J is an intrinsic physical property of w, it is equally clear that R is not an intrinsic physical relation among $P(w)$. The a priori arguments against pure physical holism would look much weaker in the face of what intuition recognizes as a concrete case of such holism. That is what quantum mechanics may be able to provide, as will be discussed in section VI below.

IV

Although philosophers are familiar with discussions of holism, the notion of nonseparability may require some introduction. In fact, there is at least one philosopher who has seen the metaphysical import of this notion, namely, David Lewis, who I believe would reject it. For the thesis of Humean supervenience that he advances in the preface to his *Philosophical Papers, Volume II*[10] comes close to formulating the contrary notion of separability. Humean supervenience is the doctrine that

> . . . all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another. (But it is no part of the thesis that these local matters are mental.) We have geometry: a system of external relations of spatiotemporal distance between points. Maybe points of spacetime itself, maybe point-sized bits of matter or aether or fields, maybe both. And at those points we have local qualities: perfectly natural intrinsic properties which need nothing bigger than a point at which to be instantiated. For short: we have an arrangement of

---

agree in their extensions in the actual world and in all naturally possible worlds, but not in *all* possible worlds. They are therefore different relations. It is a remarkable natural fact that R' holds among the $P(w)$ if and only if w exists and has J, whereas it is a trivial consequence of its definition that R holds among the $P(w)$ if and only if w exists and has J. Indeed, the former fact would be *so* remarkable as to be extremely implausible. It would be very surprising for any whole w to exist and have such an intrinsic property J irrespective of the intrinsic properties of the $P(w)$. Even when an intrinsic physical relation is included in the supervenience basis for some whole, in no ordinary case does that relation exhaust this basis. Consequently, in no ordinary case is there an intrinsic physical relation in the supervenience basis whose extension is included in that of R even in the actual world: coincidence in extension in all naturally possible worlds would be quite extraordinary. The cases we have been considering in the present paragraph are therefore quite bizarre. It is satisfying that the exclusion of R from the supervenience basis can be given a sound justification even in such bizarre cases.

[10] New York: Oxford, 1986.

> qualities. And that is all. There is no difference without difference in the arrangement of qualities. All else supervenes on that (*ibid.*, p. x).

If one restricts the scope of Humean supervenience so that it claims just that all purely *physical* phenomena supervene on perfectly natural intrinsic *physical* properties that need nothing bigger than a point at which to be instantiated, then one naturally arrives at a physical principle that may be formulated in terms of the notion of spatiotemporal separability. A physical process may be said to be *spatiotemporally separable* in *R* (where *R* is a space-time region) if and only if it is supervenient upon an assignment of qualitative, intrinsic physical properties at space-time points in *R*.

> *Spatiotemporal separability:* Any physical process occurring in a region *R* of space-time is spatiotemporally separable in *R*.

A physical process that violates the principle of spatiotemporal separability in some region may be said to be *nonseparable*. *Nonseparability* holds just in case some physical process is nonseparable. If spatiotemporal separability holds, then every physical process is determined by what happens locally—that is to say, by what is happening at each space-time point where that process is going on.

It is a natural assumption that most, if not all, ordinary physical processes are spatiotemporally separable. For example, the boiling of a kettle of water is spatiotemporally separable in a region containing the kettle over a period of time during which the water boils. The boiling of the water consists in the increased kinetic energy of its constituent molecules permitting each to overcome the short range attractive forces that otherwise hold it in the liquid. It thus supervenes on the assignment, at each space-time point on the trajectory of each molecule, of intrinsic physical properties to that molecule (such as its kinetic energy), perhaps together with intrinsic physical properties representing the magnitude and direction of the attractive force acting on that molecule at that point. The fall of a stone is similarly spatiotemporally separable in a space-time region that contains it. For the stone consists of atoms, each of which has a space-time trajectory on which various intrinsic mechanical properties are defined at each point, and each atom is acted on by various force fields, importantly including the gravitational field due to the earth, which are also defined at each point on its trajectory. It is worth noting, however, that spatiotemporal separability would appear to fail even here if one attempted to describe the falling stone without assigning a gravitational field at points of space-time. For such a description would require a specification of the distance of the stone from the center of the earth, and this is not an intrinsic physical property (or relation) holding at points confined to the region in which the stone is falling. We see here that maintaining the separability principle provides a motive for the introduction of fields, where these are taken to involve assignment of intrinsic physical properties to, or at, points of space-time independently of whether or not these points are occupied by enduring physical objects.

More generally, any physical process described fully by a local space-time theory will conform to the principle of spatiotemporal separability. For such a theory proceeds by assigning geometric objects (such as vectors or tensors) to each point in space-time to represent physical fields, and then requiring that these satisfy certain field equations.[11] But processes described fully by theories of other forms will also be spatiotemporally separable. This is true not only of pure field theories, but also of many theories that assign properties to particles at each point on their trajectories. Of familiar physical theories, it is only theories involving direct action between spatially separated particles whose models are nonseparable, and even they describe processes that are spatiotemporally separable within sufficiently large regions of space-time. It is interesting that the reformulation or replacement of such theories by field theories has often been thought of as a conceptual advance. I conjecture that concern to preserve the principle of spatiotemporal separability is one significant reason behind this perception (while another, more familiar, reason is provided by the wish to exclude action-at-a-distance).

No one has been more strongly drawn to field theories in physics than Albert Einstein. It is therefore not surprising that, in his writings on quantum mechanics, he himself formulated principles closely related to the principle of spatiotemporal separability. Immediately after offering one such formulation, he relates it to field theory.

> If one asks what, irrespective of quantum mechanics, is characteristic of the world of ideas of physics, one is first of all struck by the following: the concepts of physics relate to a real outside world, ... It is further characteristic of these physical objects that they are thought of as arranged in a space-time continuum. *An essential aspect of this arrangement of things in physics is that they lay claim, at a certain time, to an existence independent of one another, provided these objects 'are situated in different parts of space'*. . . . This principle has

---

[11] The notion of a local space-time theory is defined in John Earman and John Norton, "What Price Spacetime Substantivalism? The Hole Story," *British Journal for the Philosophy of Science*, xxxviii (1987): 515–25. It is implicit in Michael Friedman's *Foundations of Spacetime Theories* (Princeton: University Press, 1983), which formulates special and general relativity, as well as Newtonian kinematics and gravitational theory, as local space-time theories.

> been carried to extremes in the field theory by localising the elementary objects on which it is based and which exist independently of one another, as well as the elementary laws which have been postulated for it, in the infinitely small (four-dimensional) elements of space.[12]
>
> (I have emphasized the sentence in this passage which, following Howard, I take to express Einstein's separability principle.)

It may appear at first sight that the principle Einstein formulates here is concerned with spatial rather than spatiotemporal separability. But the fact that Einstein goes on to refer to four-dimensional elements of *space*, together with the fact that the phrase 'are situated in different parts of space' appears in quotes, seem to indicate that his principle is indeed to be understood as concerning spatiotemporal separation. In the next section, I shall consider another principle which may be abstracted from Einstein's writings which is concerned explicitly with *spatial* rather than spatiotemporal separability.

V

Holism and nonseparability are quite distinct notions, as is now apparent. Holism has to do with the irreducibility of certain part-whole relations, while nonseparability is to be understood in spatiotemporal terms. Why, then, have philosophers of physics confused the two?[13] This may well be because there are such subtle and interesting connections between these notions. It is these which I now propose to explore.

Recall the following statement of pure physical holism:

> *Pure physical holism:* There is some set of physical objects from a domain D subject only to processes of type P, not all of whose qualitative, intrinsic physical properties and relations are supervenient upon the qualitative, intrinsic physical properties and relations of their basic physical parts (relative to D and P).

There are two related respects in which this formulation remains problematic. The notion of an intrinsic physical relation is purely formal: it is unclear what would count as such a relation. But in the absence of such clarification, the second a priori argument against pure physical holism still has weight. Without any clear idea as to what would count as an intrinsic physical relation, how can one justify excluding the relation *R* defined in section III from that category? The following reformulation solves both problems, but at a price:

---

[12] This passage is quoted from pp. 170–1 of *The Born-Einstein Letters* (New York: Macmillan, 1971). It is a translation by Irene Born from the original German included in a letter of Einstein to Born dated 5 April 1948, in which Einstein encloses a copy of an essay that appeared as "Quantenmechanik und Wirklichkeit," in *Dialectica*, II (1948): 320–4.

[13] See, for example, Howard, "Holism, Separability and the Metaphysical Implications of the Bell Experiments," and Teller, "Relativity, Relational Holism, and the Bell Inequalities," both in J.T. Cushing and E. McMullin, eds.; and Redhead.

> *Pure spatial holism:* There is some set of physical objects from a domain *D* subject only to processes of type *P*, not all of whose qualitative, intrinsic physical properties and relations are supervenient upon the qualitative, intrinsic physical properties and the spatial relations of their basic physical parts (relative to *D* and *P*).

Instead of speaking of intrinsic physical relations among basic parts, pure spatial holism mentions only their spatial relations. Spatial relations are of special significance because they seem to yield the only clear examples of qualitative, intrinsic physical relations required in the supervenience basis in addition to the qualitative, intrinsic properties of the relata.[14] Other intrinsic relations supervene on spatial relations. This is true, for example, of the physical relation that holds between two point masses just in case they are presently attracting one another by a force of one gram weight (at least according to Newtonian gravitational theory): for this relation supervenes on the masses and their spatial separation. By restating the principle of physical holism so as to refer only to spatial relations in specifying the supervenience basis, one renders that principle more concrete, apparently without affecting its essential thrust. Moreover, since the relation *R* defined in section III is clearly not a spatial relation, one cannot argue that pure spatial holism is trivially false on the grounds that *R* must be included in that basis because it is an intrinsic physical relation. But the resulting formulation is less elegant because it treats intrinsic properties and relations differently. And it is less general, because it fails to leave open the possibility that the progress of physics may reveal additional nonsupervenient, intrinsic physical relations that should also be included in the supervenience basis.

The shift from pure physical holism to pure spatial holism has introduced some spatiotemporal relations, and has thereby linked a holist thesis more closely to nonseparability. We can forge a connecting link in the other end of the chain by introducing the notion of spatial separability.

---

[14] Teller in fact believes that spatial relations themselves supervene on intrinsic physical properties of objects (see his paper, "Space-time as a Physical Quantity," in R. Kargon and P. Achinstein, eds , *Kelvin's Baltimore Lectures and Modern Theoretical Physics* (Cambridge: MIT, 1987), pp. 425–47). If this turns out to be correct, then the reference to spatial relations in the statement of pure spatial holism will prove to be redundant.

Consider some physical process involving interactions among a number of enduring physical systems throughout some interval of time: for example, the motions of the planets around the sun. There will be some compound system (such as the solar system) of which these are all component subsystems. If the process is spatiotemporally separable, then it is supervenient upon an assignment of qualitative, intrinsic physical properties at space-time points in the region in which it occurs. Suppose that qualitative, intrinsic property assignments are made at just those space-time points within the region which lie on the trajectory of each component subsystem. (Such an assignment would suffice to specify the dynamical behavior of the solar system.) If all these assignments, together with the spatial relations obtaining between the components at each moment, are compatible with each of two or more different assignments of qualitative, intrinsic physical properties to the compound system at each moment throughout the interval, then the process is nonseparable. This kind of nonseparability results from a breakdown of what might be called *spatial separability*.

> *Spatial separability:* The qualitative, intrinsic physical properties of a compound system are supervenient on the qualitative, intrinsic physical properties of its spatially separated component systems together with the spatial relations among these component systems.

A number of Einstein's formulations of separability occur in the context of his discussions of pairs of separated systems, such as those which occur in the famous Einstein-Podolsky-Rosen paradox. Here is one such formulation:

> The real state of [the pair AB] consists precisely of the real state of A and the real state of B, which states have nothing to do with one another.[15]

I think we can recognize the spatial separability principle as a natural generalization of Einstein's formulation, if we read the last clause of the latter, not as excluding causal influence, but rather as requiring that the real states of *A* and *B* not be derivative from some more fundamental state of the compound system *AB*.[16] It is also possible to see the spatial separability principle as a generalization and correction of the principle Howard states in his article, "Einstein on Locality and Separability" (see fn. 2/3), that "any two spatially separated systems possess their own separate real states" (*op. cit.*, p. 173). Correction is necessary since Howard's principle would hold if two spatially separated systems had separate real states, even when the whole they composed had a real state that did *not* supervene on the separate real states of its components.

---

[15] This formulation appears in a letter from Einstein to Schrödinger dated June 19, 1935. I have quoted it from Howard, "Einstein on Locality and Separability," p. 180. The Einstein-Podolsky-Rosen paradox itself first appears in Einstein, B. Podolsky, and N. Rosen, "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" *Physical Review,* XLVII (1935): 777–80.

[16] In fact, Einstein himself may not have understood the principle this way. For he takes the following locality condition (which appears in his next sentence) to follow immediately from his spatial separability principle, which on my reading it does not, without additional causal assumptions. "The real state of B . . . cannot

Spatiotemporal separability entails spatial separability. The argument is as follows. Suppose that spatial separability fails. Then there is a time *t* at which some compound system has a qualitative, intrinsic physical property *P* that does not supervene upon the qualitative intrinsic properties and relations of its spatially separated components. If some of these components are localized at spatial points, then their qualitative, intrinsic physical properties will generate assignments of qualitative, intrinsic physical properties at the space-time points defined by these spatial points at *t*. There may be other ways in which the qualitative, intrinsic physical properties and relations of the components restrict the compatible assignments of qualitative, intrinsic physical properties at space-time points occurring at *t*. Let R_t be the space-time region defined by a time-slice at *t*. Consider an assignment *A* of qualitative, intrinsic physical properties at space-time points in R_t which is compatible with the qualitative, intrinsic physical properties and relations of the components at *t*. If spatiotemporal separability held, then a process involving the compound system at *t* would have to be separable in R_t. But, by assumption, there is such a process that involves the compound system's having property *P* at *t*. Hence assignment *A* would have to determine that the compound system has *P* at *t*. But *A* was any assignment compatible with the qualitative, intrinsic physical properties and relations of the components at *t*. And so the qualitative, intrinsic physical properties and relations of the components at *t* would have to determine that the compound system has *P* at *t*. This would contradict the earlier conclusion that *P* does not supervene upon the quali-

---

depend on the kind of measurement I carry out on A." Note that, if one takes the real state of a system to be given by its quantum-state vector (as does one popular version of the Copenhagen interpretation), then quantum mechanics immediately conflicts with Einstein's spatial separability principle. For the singlet quantum state of a system *AB* composed of two subsystems *A* and *B*, each of nonzero spin, does not consist of the quantum state of *A* and the quantum state of *B*. But Einstein, for one, did *not* take the real state of a system to be given by its quantum-state vector. Indeed, one of his strongest reasons for denying that the quantum state of a system gives its real state, and consequently for rejecting this version of the Copenhagen interpretation, was his desire to interpret quantum mechanics in a way consistent with his spatial separability principle.

tative, intrinsic physical properties and relations of its components at *t*. Therefore, spatiotemporal separability fails. Contrapositively, spatiotemporal separability entails spatial separability.

Spatial separability is therefore linked to spatiotemporal separability because it is entailed by it. But spatial separability is also closely connected to spatial holism. To make this clear, consider the following denial of spatial separability:

> *Spatial nonseparability:* There exists a compound physical system, not all of whose qualitative, intrinsic physical properties supervene on the qualitative, intrinsic physical properties of its spatially separated component systems together with the spatial relations among these component systems.

Compare this with the principle of spatial holism stated earlier.

> *Pure spatial holism:* There is some set of physical objects from a domain *D* subject only to processes of type *P*, not all of whose qualitative, intrinsic physical properties and relations are supervenient upon the qualitative, intrinsic physical properties and the spatial relations of their basic physical parts (relative to *D* and *P*).

These two principles are clearly very similar, even though I have taken one to express a form of holism, and the other to express a form of nonseparability. With inchoate versions of these last two principles in mind, it would be easy to blur the distinction between holism and nonseparability, and to regard these distinct notions as essentially equivalent.

On one natural supposition, pure spatial holism entails spatial nonseparability. For suppose that pure spatial holism obtains because there is a qualitative, intrinsic physical property *P* of some compound system which does not supervene on the qualitative, intrinsic physical properties and spatial relations of its basic physical parts at some time when these parts are all spatially separated. In that case it follows that *P* also fails to supervene on the qualitative, intrinsic physical properties of these spatially separated component systems together with the spatial relations among them, and so spatial nonseparability obtains. (It might not be enough for the basic parts merely to *compose* the spatially separated components, for if they did so in such a way that the properties of any spatially separated component failed to supervene on the properties of its basic parts, then pure physical holism might hold while spatial nonseparability failed.)

If spatiotemporal separability entails spatial separability, then spatial nonseparability entails nonseparability. And if pure spatial holism also entails spatial nonseparability, it then follows that pure spatial holism entails nonseparability. Moreover, provided that spatial relations are included among qualitative, intrinsic physical relations, pure physical holism entails pure spatial holism, in which case it further follows that pure physical holism entails nonseparability. The foregoing arguments license one to detach the consequents of these conditionals subject only to the (rather weak) provisos these arguments required.

But it is quite clear that nonseparability does not entail pure physical holism. A process may be nonseparable even though it involves only objects with no nontrivial components (i.e., objects that are elementary parts of any object that contains them). The passage of an electron through the famous two-slit experiment may well provide an instance of such a process; for there is no reason to suppose that an electron has nonzero spatial extension, or any nontrivial component parts, whereas there is some reason to maintain that an electron passes through both slits of this experiment at once, and so has a nonseparable trajectory.[17] Perhaps, then, quantum mechanics furnishes examples of nonseparability without holism. In the final section, I consider whether it provides examples of nonseparability arising from an underlying holism.

VI

It follows from the foregoing analysis that, in order to determine whether or not quantum mechanics establishes either pure physical holism or nonseparability, it is necessary first to decide what are the qualitative, intrinsic physical properties and relations of quantum systems. This is by no means a straightforward matter. It may be easier to begin with the corresponding question about classical systems.

Classical particle mechanics ascribes two kinds of physical properties to the systems with which it deals. On the one hand, there are properties (such as a particular value of mass or charge) possessed by every system of particles of a given type, which remain unchanged as long as that system exists: we may call these the *fixed properties* of the system. On the other hand, there are properties associated with quantities called *dynamical variables* (such as energy; component of position, of momentum, of angular momentum). A system of a given type may have one of these *dynamical properties* at one time while failing to have it at some other time. For any "nice" set of real numbers, and any dynamical variable pertaining to a system, there is a dynamical property that the system possesses just in case the value of that variable on the system is restricted to that set. In particular, for each unit set, there is a dynamical property that the system has just in case the variable has the value contained in that set: we may call this a *precise value property*.

---

[17] The key feature of such examples is the attribution to a quantum system of nonlocalized spatial properties, such as an interval-valued (rather than a real-valued) position; cf. my *The Philosophy of Quantum Mechanics: An Interactive Interpretation* (New York: Cambridge, 1989), which offers an interpretation of quantum mechanics with this feature. See also Teller, "Quantum Mechanics and the Nature of Continuous Physical Quantities," this JOURNAL, LXXVI, 7 (July 1979): 345–61.

The possible dynamical properties of a system are generally taken to be determined by the structure of an abstract mathematical space (a phase-space) associated with systems of that type. The phase-space for a single particle is six-dimensional: each point in the space represents a precise real value for each of the three components of the particle's position, and each of the three components of its momentum. The phase-space for a compound system composed of N particles is $6N$–dimensional, and each point represents the position and momentum of every component particle. In classical particle mechanics, the instantaneous dynamical state of a system (whether simple or compound) is given by the assignment of a precise value property to each dynamical variable. It may be completely specified by assigning to the system a point in phase-space. Since all dynamical variables are functions of position and momentum, this assignment generates an ascription of a precise real value to every dynamical variable pertaining to that system. Each dynamical property of a system is associated with a "nice" subset of phase-space: the system has that property if and only if its dynamical state is represented by a point lying in that subset of phase-space.

It would be natural to assume that all fixed properties of a classical particle system are among its qualitative, intrinsic physical properties, as are all dynamical properties corresponding to any "nice" region of phase-space. One might also assume that these comprise the qualitative, intrinsic physical properties of a classical particle system, and that such systems have no qualitative, intrinsic physical relations that do not supervene on these. But both these assumptions may be questioned.

Among the dynamical properties of a classical particle are precise value properties associated with components of position. A property such as $x = 3$ ascribing a precise x-coordinate to the particle is not itself an intrinsic property of the particle, however. Rather, it corresponds to a spatial relation between that particle and an (arbitrarily chosen) origin of spatial coordinates, namely, the relation of being 3 units (in some system) distant from the origin along a certain spatial direction. That relation is actually an intrinsic relation involving the particle: its other relata may be particles, points of space or space-time, or a combination of these. Moreover, it does not seem to supervene on intrinsic physical properties of its relata.[18] If it is necessary to give an intrinsic specification of the dynamical state of the particle which implies the ascription to it of a property such as $x = 3$, then intrinsic spatial relations involving that particle must be included in the specification. But one might wonder whether that is necessary. For a one particle system is highly idealized, and one might suspect that, for a multiparticle system, the only spatial properties that need to be included in an intrinsic specification of its dynamical state are those connected to the spatial relations among just these particles.

Consider then a two-particle classical system. One dynamical property represented by a region of the 2-particle phase-space corresponds to the distance between the two particles being exactly 3 centimeters. As well as being a property of the 2-particle system, however, this also corresponds to a relation between the two particles that compose that system. Moreover, this is a qualitative, intrinsic physical relation between these two particles which may well not supervene on their qualitative, intrinsic physical properties. The phase-space representation of the state of a compound system does not represent qualitative, intrinsic physical relations among that system's components as such, and may therefore mislead one into overlooking some such relations. But in classical mechanics one need not employ this representation. It would be more perspicuous for present purposes to specify the dynamical state of a set of classical particles by giving their relative positions, in addition to their individual momenta, at each instant.[19]

Quantum mechanics also ascribes to its systems two kinds of qualitative, intrinsic physical properties, many of which are familiar from classical particle mechanics.[20] There are fixed properties, now also including intrinsic spin. And there are dynamical properties, including components of spin. In quantum mechanics, the dynamical properties of a system, whether simple or compound, are associated not with subsets of a phase-space, but with subspaces of a different kind of abstract mathematical space (a Hilbert space) appropriate to systems of that type.

---

[18] But cf. Teller, who proposes in his "Space-time as a Physical Quantity" that such spatial relations are indeed supervenient on qualitative, intrinsic physical properties of objects.

[19] One might think it unnecessary to give the momenta in addition, on the grounds that momentum is derivative on mass and velocity, and relative velocity may be derived from change in relative position. But classical dynamics also requires a specification of absolute acceleration that cannot be derived from relative velocities or their changes This raises the interesting question as to whether further qualitative, intrinsic physical relations underlie the specification of momenta—perhaps including a relation that holds between a particle and points of space-time just in case that particle's trajectory is inertial.

[20] I restrict attention to the nonrelativistic quantum mechanics of a fixed set of particles. The case of quantum field theories raises further difficulties concerning

The instantaneous state of a system is not given in quantum mechanics by assigning a precise real value to every dynamical variable pertaining to that system. Rather, the state of a system at a time may be described by assigning to it a mathematical object of a certain kind—a wave function or, more generally, a *state vector* (which is an element of the Hilbert space associated with the system). For a system described by a given state vector at a particular time, quantum mechanics contains an algorithm that specifies, for each dynamical property, the probability that an observation of that property would reveal it to be possessed at that time. The algorithm is such that this probability is one just in case the state vector lies in the subspace associated with that property. It follows that there is no state vector that assigns probability one to some precise value property associated with each dynamical variable. In particular, no state vector assigns probability one to a precise value property associated with a given component of position and also to a precise value property associated with the same component of momentum.

What is the connection between the state vector of a system at a time and the dynamical properties it possesses at that time? Different interpretations of quantum mechanics answer this question differently. According to one popular version of the Copenhagen interpretation, a system has a dynamical property at a time if and only if it may be described by a state vector that lies in the subspace associated with that property. It is an immediate consequence of this version that no system ever has simultaneous precise values of both position and momentum. But on other interpretations one cannot "read off" a system's dynamical properties from its state vector in this way: the specification of a system's dynamical properties requires additional, or other, information. Call the set of dynamical properties possessed by a system at a time its *dynamical state* at that time. While interpretations differ as to whether or not a system's dynamical state is determined by a state vector associated with the

---

the character of the systems these theories may be taken to describe. In order to decide whether or not such theories manifest holism or nonseparability it will therefore be necessary, first, to resolve disputed ontological questions raised by quantum field theory system, on no interpretation does quantum mechanics itself prescribe the dynamical state except via this state vector.[21]

What, then, are the qualitative, intrinsic physical properties and relations of systems described by quantum mechanics? These clearly include a system's fixed properties. Presumably, they also correspond in some way to its dynamical properties, whether or not these are determined by any state vector. Perhaps a system's state vector should also be included among its qualitative, intrinsic physical properties. Quantum systems may even have further qualitative, intrinsic physical properties or relations that quantum mechanics does not describe at all (corresponding to so-called "hidden variables"). A decision as to whether or not quantum mechanics demonstrates nonseparability or physical holism requires a definitive characterization of the qualitative intrinsic properties and relations of quantum systems. But such a characterization cannot be given without resolving problems central to the interpretation of quantum mechanics. Nevertheless, one can show that there are significant interpretations of quantum mechanics according to which this theory does indeed demonstrate nonseparability, and arguably also pure physical holism.[22] Perhaps the easiest way of doing this requires a preliminary discussion of an important class of compound quantum systems, namely, spin-singlet pairs.

A pair of quantum systems *A*, *B*, each with spin ½, may be in a spin-singlet state. They then constitute a compound quantum system *AB* with certain properties, such as having total spin zero; and each of *A*, *B* separately is a basic part of *AB* relative to the domain *D* of pairs composed of subsystems of these types, and any type *P* of process through which systems of these types are preserved. (If, for example, *A* is an electron and *B* is a positron, then each may well be an ultimate part of *AB*, and in any case each is a basic part of *AB* relative to processes that do not create or annihilate electrons or positrons.) As long as *AB* remains in the spin-singlet state, it retains the dynamical property of having total spin zero. One fixed property of *AB* is that its electric charge is the sum of the electric charges of *A* and of *B*. Another property of *AB* is that it is described by a state vector that specifies a characteristic set of probabilistic correlations for joint measurements of spin components, one on *A*, the other on *B*. If, for example, the spin components along some one direction of *A* and of *B* are simultaneously measured, the probability that each has the same value is zero, while the probability that the two spin components have opposite values is one. Collectively, the spin-singlet correlations violate the famous Bell inequalities—a fact of some significance, since it rules out certain potential explanations of these correlations, as well as certain potential interpretations of quantum mechanics.[23]

---

[21] It is sometimes convenient to describe a system's state not by a vector, but by a more general mathematical object (e.g., a ray, or a density operator) Such descriptions are generally understood to be either equivalent to, or less detailed than, descriptions in terms of a state vector. While there are theories that provide a more detailed specification of the dynamical state of a system, these theories do not claim that quantum mechanics itself yields this specification. Rather, they supplement the quantum mechanical description, typically by the addition of so-called "hidden variables" whose values are not determined by the state vector. Bohm presents one such theory in his "A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden Variables', I and II," *Physical Review*, LXXXV (1952): 166–93.

[22] Since my main aim in this paper is to clarify, not to settle, the issues of quantum holism and nonseparability, it would be inappropriate for me to assume or defend any particular interpretation of quantum mechanics here. In my *The Philosophy of Quantum Mechanics: An Interactive Interpretation*, I propose and defend an interpretation that is similar in presently relevant respects to that version of the Copenhagen interpretation whose bearing on holism and nonseparability I consider below.

On the version of the Copenhagen interpretation mentioned above, a spin-singlet pair *A*, *B* clearly establishes nonseparability and arguably also pure physical holism. *A*, *B* are the basic parts of the compound system *AB*. On this view, the qualitative, intrinsic physical properties and relations of *A* correspond to just those fixed and dynamical properties recognized by quantum mechanics, and similarly for *B*. Since neither *A* nor *B* has a state vector in the singlet state, this cannot be included as a further intrinsic property of either.

Just as in classical particle mechanics, the instantaneous dynamical properties of *A*, *B*, and *AB* do not directly reveal any qualitative, intrinsic physical relations between *A* and *B*. But this does not imply that there are no dynamical relations between them. Some of these supervene on the qualitative, intrinsic physical properties of *A* and of *B*. Just as in classical particle mechanics, there may be other, nonsupervenient, concealed intrinsic dynamical relations, including spatial relations. While in classical particle mechanics these may be revealed by moving from a phase-space representation to a direct specification of the relative spatial positions of *A* and *B* (together with other of their dynamical properties such as momentum), on the interpretation of quantum mechanics presently under consideration, however, this is not so straightforward. On this interpretation, specifications of the dynamical states of *A* and *B* cannot be taken to include a spatial relation corresponding to a precise value for the distance between *A* and *B*.[24] On the other hand, they do include "weaker" spatial relations, such as "not being spatially coincident" or, perhaps, "being at least ten meters apart." Thus, while there are intrinsic spatial relations between *A* and *B*, these cannot be derived from a single fundamental relation of being separated by a certain definite distance.

---

[23] Including local, hidden variable theories and perhaps also Einstein's interpretation Whether Einstein's interpretation is, in fact, ruled out on these grounds remains controversial.

Nonseparability follows immediately. For assume that *A* and *B* are ultimate, and so the only nontrivial, parts of *AB*, and that no other objects figure in some process involving the total spin of *AB*. Since *A* and *B* are not spatially coincident, this process cannot supervene on an ascription of qualitative, intrinsic physical properties at their common space-time locations. But if it is to supervene on any ascription of qualitative, intrinsic physical properties at space-time points, then these must be ascriptions of properties to *A*, *B*, or *AB*. Hence, no such process can be supervenient upon an ascription of qualitative, intrinsic physical properties at space-time points where that process is occurring.

Pure physical holism may be argued to hold as follows. On this interpretation, intrinsic physical properties of *AB* like total spin do not supervene even on the qualitative, intrinsic physical properties and relations of *A*, *B* corresponding to those which figure in the supervenience basis for the corresponding classical system, suitably enriched by the new qualitative, intrinsic physical properties of *A*, *B* associated with the nonclassical quantity intrinsic spin. Since the move from classical to quantum mechanics otherwise weakens the supervenience basis (for example, by removing a relation corresponding to precise relative positions of *A*, *B*), these properties of *AB* clearly also fail to supervene on the qualitative, intrinsic physical properties and relations of *A*, *B* postulated by quantum mechanics.

For example, *AB* has total spin zero. That *AB* has this property follows, on the present interpretation, from the fact that its state vector in the spin-singlet state lies in the subspace corresponding to this property (and, consequently, a measurement of total spin will, with probability one, reveal that it is zero). But *AB*'s possession of this property does not supervene on qualitative, intrinsic physical properties and relations of *A* and *B* separately. In particular, it does not supervene on the individual spins of *A* and of *B*, since (on this interpretation) *A* and *B* have no nontrivial intrinsic spin properties in the singlet state. *A*, *B* are basic parts of *AB*, relative to any class of processes that preserve systems of these types. Consequently, we have an instance of pure physical holism.

---

[24] A necessary condition for it to include such a relation is that the quantum-state vector of the joint system *AB* lie in the subspace corresponding to some precise value for the difference in coordinates of *A* and *B* along some direction. The wave function introduced in the famous paper by Einstein, Podolsky, and Rosen may seem to satisfy this condition. But, in fact, it does not, since the relevant subspace in this case actually fails to contain any nonzero vectors.

It might be objected against this argument that *AB*'s having spin zero is in fact supervenient on a certain qualitative, intrinsic physical relation *S* between *A* and *B* which might be described as "having equal but oppositely directed spins." Now this description of the supposed relation is unfortunate, since on the present interpretation neither *A* nor *B* has any definite spin in the singlet state. But perhaps quantum mechanics should be understood to postulate such an intrinsic physical relation, however it is referred to? If so, then no instance of pure physical holism has been presented, since *AB*'s having spin zero is determined by the fact that *A*, *B* bear one another this relation.

Notice that this objection is closely related to the second a priori argument against pure physical holism considered and rejected in section III. That the supposed relation *S* holds between two objects would not suffice for the pair *AB* to exist. But, given that the pair *AB* exists, this relation's holding between *A* and *B* would be necessary and sufficient for *AB* to possess the property *J* of "having spin zero."

But, in fact, there are reasons for denying that *S* is an intrinsic physical relation postulated by quantum mechanics. It is not an accident that, while "having spin zero" is an important quantum mechanical property of the pair *AB*, the relation *S* is not considered significant enough to warrant a name. The total spin of a system is a physical property that is fundamental to the behavior of that system in quantum mechanics. Its spin is determined by the spins of its component systems, together with a way of combining these spins (the vector addition law). But the property so determined is of the very same kind as the properties that help to determine it. As far as its spin goes, it is irrelevant whether or not a system is composed of subsystems: quantum mechanics applies to the spin of a system in just the same way in either case. This is important, since it permits one to treat the total spin of a complex system like a silver atom just as one would that of a spin ½ system with no nontrivial subsystems.

These considerations provide weighty reasons why total spin is an intrinsic physical property of a quantum system, whether that system is simple or compound. Of course, given that a pair of spin ½ systems has total spin zero, it follows that the systems in that pair are related by the physical relation *S*. But there is no reason to believe that *S* is an intrinsic rather than a derivative physical relation.[25] This answers the objection to the claim that such pairs establish pure physical holism on the interpretation of quantum mechanics presently under consideration.

It is worth pointing out that there are interpretations according to which quantum mechanics demonstrates neither nonseparability nor pure physical holism. These include interpretations of quantum mechanics involving nonlocal, hidden variables. Such interpretations have not been widely accepted. But here we have a clear parallel with the status of indeterminism in quantum mechanics. For a nonlocal, hidden variable interpretation of quantum mechanics might, if accepted, render quantum mechanics compatible also with determinism. Nevertheless, the impact of quantum mechanics has prompted contemporary metaphysics to accept the possibility of physical indeterminism. It must similarly come to terms with the possibility of physical holism and nonseparability. We could know for sure whether or not quantum mechanics demonstrates either nonseparability or pure physical holism only if we both understood and believed quantum mechanics. Metaphysics can and must proceed without this knowledge.

RICHARD A. HEALEY

University of Arizona

---

[25] Notice that *S* counts as an external relation on Lewis's classification (see *On the Plurality of Worlds*, p. 62). This suggests that he would classify *S* along with spatial relations as an intrinsic relation. But this points to a difficulty with his scheme. Lewis divides properties into intrinsic and extrinsic based on whether things have them in virtue of the way they themselves are, or in virtue of their relations or lack of relations to other things. An analogous division among relations would count spatial relations, but not *S*, as intrinsic. For *A* and *B* have *S* in virtue of the way they are related to *AB*, rather than in virtue of the way they are related to one another.

BOOK REVIEWS

Sources of the Self. CHARLES TAYLOR. Cambridge: Harvard University Press, 1989. xii, 601 p. Cloth $37.50.*

Modern moral philosophy, Charles Taylor tells us, has suppressed or "occluded" discussion of the human good. Since he finds "[t]he very claim not to be oriented by a notion of the good" to be "incredible" (489), Taylor sees this as a failure. Ignoring the work of G. E. Moore, Max Scheler, Nicolai Hartmann, R. B. Perry, and other value theorists, he offers a remarkably original approach to the topic. The self, on Taylor's view, is not an objectively given entity like a star or a bacillus. Its identity is constituted largely by its values. To study the good is thus to study what makes us the selves we are. And we cannot study the sources of the self statically: only a historical narrative will enable us to understand the variety of goods that provide the ultimate rationales for our present ways of life and thought.

Taylor's study of the good thus goes beyond mere value theory. It is shaped to continue his long-standing battle against reductivist and naturalist epistemologies and ontologies generally. As he has explained elsewhere, he thinks the scientific and epistemological arguments for such views are extremely weak. They really draw their appeal from unavowed moral grounds. Other critics, from Nietzsche through Michel Foucault, have tried to show this, but their own positions are "often underdemonstrated, indeed rather impressionistically argued for." Moreover, Taylor thinks that we cannot just dismiss the naturalistic view of the self. It is "far from being simply wrong" and we are all "deeply imbued with it." We must rather try to see its limits by an examination of "the spiritual roots of naturalism" which shows "how, through the whole course of the development of the modern identity, the moral motivation has been intertwined with the epistemological." The latter by itself would never have been sufficient to win the support that reductionism and naturalism now have. "What would ultimately carry conviction" for his view, Taylor claims, "would be an account of this development which illuminated it and made more sense of it than its rivals, and particularly than naturalistic ones. As for any hermeneutic explanation," he adds, "interpretive plausibility is the ultimate criterion."[1]

---

* My thanks to Edward Minar for helpful comments on a draft.

[1] *Human Agency and Language: Philosophical Papers,* Vol. I (New York: Cambridge, 1958), pp. 6–8. The whole "Introduction" and the first two papers in this volume form an almost indispensable prolegomenon to *Sources.*

0022-362X/91/8808/422–6

This helps us to see both the direction of Taylor's complex discussion in *Sources* and the nature of his argument. His narrative portrays the values of modernity evolving in interaction with its epistemologies and ontologies, each supporting and being supported by the other. He expects us to find that his history gives us a better understanding of the modern self than any explanation a reductionist can give. We shall then conclude that Taylor's account of the role of values in our lives is preferable to any account available to such theories. And this in turn will support Taylor's account of the domain of value.

No mean enterprise, this, and made more difficult by Taylor's own predilection for a theistic understanding of the ultimate good. St. Augustine, addressing God at the opening of the *Confessions,* said: "you made us for yourself and our hearts find no peace until they rest in you." Hobbes took our perpetual restlessness to be simply an indication of our being alive, not a result of mistaking earthly goods for that heavenly good that alone gives peace. Taylor is on St. Augustine's side. Although accepting many of the values of modernity, he thinks them in an important way inadequate: they impose demands without providing the motivational resources that would make compliance possible. All of us, apparently, would be dissatisfied if we would reflect on our situation. Taylor hints that only a return to a theistic moral source provides a way out.

In Part I, Taylor explains his idea of the self and of the "strong evaluations" that are necessary to it. Although our values are discovered through feelings, only a bad epistemology will suppose that this entails their entire subjectivity: our feelings reveal aspects of the world that are truly there for us (257). The "incomparably higher" goods thus made available to us enable us to assess our own desires and to find point and meaning in our ways of life and thought. We often are not aware of what these goods are, since articulating them requires work of a kind that reductionist philosophy discourages.[2] Presumably a satisfactory articulation is one that generates a hermeneutically plausible narrative. From here on, therefore, we get history.

Taylor discusses classical theorists (mainly Plato) and Augustine, but he is primarily concerned with developments from the end of the sixteenth century onward. The ancients are included to show that there was a time when sources of value were seen as external to the self. Augustine made a turn inward and since then values have been found increasingly by delving within the self, at first going through the self to find a source outside it (Taylor points to Descartes's proof of God's existence from the idea of God in his own mind) and ultimately finding the sources of value within the self as such. Two main lines of thought interest Taylor. Descartes and Locke moved toward seeing the self as a disengaged entity capable of controlling desire and of viewing everything other than itself as having at most instrumental value. The self is the autonomous chooser whose choices bestow value on everything else. The second line became prominent with Rousseau's rejection of this detached self and with the emergence of the Romantic expressive self. The self thus understood is first seen as the source of value when it is truly in contact with nature and then, more radically, as the source when it is expressing its own nature or constructing, through language, what it also expresses.

---

[2] 'Articulation' is a key term; for more explanation than is given in *Sources,* see "Self-Interpreting Animals" in the volume noted above.

What gives these versions of the self their appeal, Taylor holds, is that they seem to support or validate ways of living that have come to be especially valued in modern times. We believe in self-control and self-governance; we think increasingly of all people as equally deserving of respect; we value common life, not the elite lives of warrior heros, political leaders, or monks; we try to reduce suffering and pain, and so we demand of ourselves that we be benevolent always to everyone. If God was still a moral source for Descartes and Locke—if belief in God enabled them to see these values as motivating them to action—the tendency they represent toward finding sources within the self has now gone to extremes. The modern predicament is that the theories of the self we have accepted as supporting these values are coming to seem incapable of providing the moral sources of motivation we need in order to live as the values dictate. Taylor develops this thought largely through discussions not of philosophers but of post-Romantic literature.

For Taylor, then, the problem is not the one Alasdair MacIntyre described in *After Virtue*,[3] that we do not agree on common daily values. There are tensions between instrumentalists and expressivists, Taylor says, but all agree in placing high value on individual freedom, reduction of suffering, and universal benevolence and justice. This consensus makes our rare serious disagreements (for example, over abortion) stand out sharply (496). But our basic problem is that we cannot locate an incomparably higher good, through which we can come to take full satisfaction in the life goods to which we are committed (515ff). Such a good cannot be purely external: it must resonate within the subject, while not being merely subjective (510). It is presumably this need to find an incomparably higher good uniting us with an objective order which might be satisfied by a return to a theistic source; but Taylor barely suggests this (342, 521).

---

[3] Notre Dame: University Press, 1981.

Sources is a rich and rambling book, provocative and often illuminating as history, and philosophically interesting in showing us how Taylor's antinaturalist program works out in the domain of value. In the space available, I can only suggest a few questions.

First, Taylor talks as if the sole alternative to his view were a subjectivistic reductionism. But we might hold that, in the course of coping with practical problems, communities have developed multifaceted languages for use in a wide variety of contexts. If the languages of evaluation serve purposes that differ from those served by the languages of science, as Taylor agrees they do, then distinctions of subjective and objective can be sustained within the discourses of value, and need not be explained by some subsistent realm of the kind Taylor seems to suggest. On such a view, moreover, there need be no single best account of value. There would be numerous accounts whose goodness would depend on what one wanted the accounts to do.

Next, Taylor has difficulty making his historical account carry the philosophical weight he loads onto it. History is to show that constitutive goods disclosed through feeling have been the crucial factor in creating the distinctively modern identity. Even allowing that explanations invoking values may be illuminating although incomplete (and Taylor carefully points out their incompleteness), they must find some actual purchase in historical reality. Taylor, however, seems so certain that thought and action must ultimately be explained in terms of visions of the good that he does little to show either that major thinkers were themselves moved by values as he says they were, or that their ideas were widely accepted because of their ties with the values Taylor thinks they involved. For instance, he does not try to show that Locke's epistemology was seen by readers as being so weak that their acceptance of it has to be explained by something other than their belief in its strength as a theory of knowledge.

This is perhaps a small point; but Taylor's a priori assurance gives rise to a more serious defect in his narrative. He unaccountably neglects the question of why discussion of the ***summum bonum*** or of substantive goods was sidelined by many of the most influential modern philosophers in the first place. He provides no sustained treatment of modern natural-law theory, the work of writers from Grotius through Kant who tried to develop explanations of morality showing how people could live together despite deep disagreements about the good. Whether reductionists or not, they thought they could construct viable programs for social life precisely by ignoring the issues Taylor takes as central. The moral psychology to support this, which emerged most explicitly in Richard Price, Thomas Reid, and Kant, is that there is a previously unappreciated complexity to our motivation. We can be moved, according to the new view, not only by our grasp of the good but by our independent understanding of the right as well. This position, which responded to pressing practical problems, constitutes one of the major innovations of modern moral thought, so that the absence of any serious engagement with it leaves a chasm in Taylor's historical argument. Even if he is right, as I believe he is, in thinking that insight into practical principles helped to shape the epistemologies and ontologies of modernity, he has not shown that it was primarily principles about the good that had this force.

Third, from Taylor's Frankfurtian view about how human identity involves second-order desires capable of controlling first-order desires, nothing follows about any need for highest-order desires—for "ultimate" sources of "incomparably higher" value to give force to all the rest. Taylor presumably wants some such principle in order to move us from the values he and we share as modern selves to something we can only find in a theistic source of moral motivation. But he gives no account of what entitles something to be a source of value not needing any further backing. He thinks that, although the truth of the theistic position is often denied, "no one doubts that those who embrace it will find a fully adequate moral source in it" (317). Even if this astonishing remark were true (and it seems to ignore at least the many dimensions of the problem of evil), Taylor has done nothing to block the question of what incomparably higher value makes a theistic source adequate. Surely, if such a question can be raised about the adequacy of the values of human dignity, autonomy, and freedom from suffering, it can be raised about a theistic source, too.

J. B SCHNEEWIND

Johns Hopkins University

God, Time, and Knowledge. WILLIAM HASKER. Ithaca: Cornell University Press, 1989. xi, 209 p. Cloth $24.95. And *The Nature of God: An Inquiry into Divine Attributes.* EDWARD R. WIERENGA. Ithaca: Cornell University Press, 1989. ix, 238 p. Cloth $28.50.

William Hasker explores God's knowledge of the future. He is particularly interested in the relationship between such knowledge and human freedom. This is a clear book, filled with subtle and provocative argumentation. Hasker provides an argument against the possibility of "middle knowledge" and hence he rejects "Molinism." He offers an ingenious argument that "simple divine foreknowledge" has no point—it would be useless with respect to God's providential activity. Hasker also argues that various responses to incompatibilism about God's foreknowledge and human freedom (associated with Aquinas and Ockham) are inadequate. Finally, Hasker offers an alternative picture according to which the future is "open" in the sense that statements about future actions need not be true or false now. Although God knows the *possibilities* of the future and also the likelihoods that these will be realized, he cannot know about future free actions (insofar as there are no truths here to be known).

In this review, I cannot undertake to discuss various of Hasker's arguments. In particular, I shall not discuss his argument against the possibility of middle knowledge.[1] Also, I shall not discuss Hasker's claim that there is no coherent difference between "counterfactual power over the past" and "bringing about the past," and thus that a certain sort of compatibilism is incoherent.[2] I shall take issue instead with the following claims of Hasker: (1) the compatibilist must implicitly be operating with the notion of a "general power" rather than "the power to perform a particular act under given circumstances," and thus the compatibilist and incompatibilist are talking past each other; (2) the compatibilist cannot coherently affirm "lib-

---

[1] For a discussion of Molinism, see Alfred J. Freddoso, *On Divine Foreknowledge (Part IV of the Concordia)*, Luis de Molina, trans (Ithaca: Cornell, 1988). Also, see the exchange between Hasker and Thomas P. Flint: Flint, "Hasker's *God, Time, and Knowledge*," *Philosophical Studies*, LX (1990): 103–16, and Hasker, "Response to Thomas Flint," *Philosophical Studies*, LX (1990): 117–26.

[2] I criticize this claim in "Introduction. God and Freedom," in my, ed. *God, Foreknowledge, and Freedom* (Stanford: University Press, 1989), pp. 18–23. For a discussion of this issue in the context of the argument for the incompatibility of causal determinism and freedom, see Carl Ginet, "In Defense of Incompatibilism," *Philosophical Studies*, XLIV (1983): 391–400; and my "Freedom and Miracles," *Nous*, XXII (1988) 235–52, esp. no 5, p. 251.

0022-362X/91/8808/427–33

ertarian free will"—the claim that we sometimes have genuine alternative possibilities of choice and action.

I shall first lay out some of the considerations in virtue of which Hasker concludes that the compatibilist about God's foreknowledge and human freedom must be operating with the notion of a "general power." Note that one can have a generalized power while lacking the ability to exercise this power: I may be able (in the general sense) to play the piano but be temporarily paralyzed or tied down, and so forth. It would be highly damaging to compatibilism if it required this general notion of a power, since it is widely acknowledged that it is the notion of the power to perform a particular act in given circumstances (rather than the general power) that is relevant (in a straightforward way) to free will (and moral responsibility).

Hasker alludes to the distinction between hard and soft facts about times. He points out that there are two "moments or aspects" of the notion of a hard fact: first, a hard fact is "really about the past," and second, if a proposition expresses a hard fact, then it cannot be in anyone's power (at the relevant time) to render that proposition false (82).[3] Although Hasker describes these two "moments or aspects" of the notion of a hard fact, he *defines* hard facts as "true propositions, such that, with respect to a given time, it is impossible that anyone at or after that time should have the power to render them false." Given this definition of hard facts, Hasker explicates hard facts in terms of the notion of "future-indifference" and a number of other analytical hypotheses (84ff).

Having explicated the notion of a hard fact, Hasker argues that facts about God's prior beliefs about future actions are *hard* facts about the relevant past times (91–5). This turns out to be the case, on Hasker's approach, even though God's beliefs are not *temporally indifferent (nonrelational)* facts; since fixity and temporal indifference are different moments or aspects of hardness, a fact can be fixed even though it is not temporally indifferent.[4]

---

[3] I do not believe that "really about the past" captures the relevant temporal aspect, because a fact can be really about the past without being *solely* about the past: the conjunctive fact, "Jack sits at T_1 and Molly stands at T_3" is really (but not solely) about the past with respect to T_2. Thus, the notion of being really about the past does not capture the idea of being "fully accomplished and over-and-done-with," which is the appropriate notion.

[4] In a previous paper, Hasker distinguished two kinds of facts that correspond to the two moments or aspects of fixity and temporal nonrelationality: "Hard Facts and Theological Fatalism," *Nous*, XXII (1988): 419–36, repr. in Fischer, ed. This previous approach is, in my judgment, much preferable insofar as confusions are engendered by the "aspects" approach. For example, having defined hard facts as those which cannot be rendered false, Hasker says, "But even if the

Consider now Hasker's definition of the power to perform a particular act under given circumstances:

> In general, if it is in N's power at *T* to perform *A*, then there is nothing in the circumstances [fn. 18: It will be recalled that the circumstances that obtain at *T* comprise all and only the hard facts with respect to T] that obtain at *T* which ***prevents*** or ***precludes*** N's performing *A* at *T*. Here 'prevent' applies especially to circumstances that are ***causally*** incompatible with N's performing *A* at *T*, and 'preclude' to circumstances that are ***logically*** incompatible with *N*'s doing so (134).

On the basis of this account and his argument that God's beliefs are hard facts, Hasker concludes that the compatibilist must have in mind the general notion of power (and thus that the compatibilist and incompatibilist are talking past each other).

But Hasker's argument here is clearly unfair. He has produced an argument for the conclusion that God's beliefs are hard facts, but it is unfair to suppose that the compatibilist accepts this argument (or even knows of it). Indeed, the argument is controversial,[5] and its conclusion is precisely what is at issue in the debate between compatibilists and incompatibilists. Thus, the claim that God's beliefs are hard cannot simply be taken for granted when one is trying to see what notion of power the compatibilist is employing. Given this point, the compatibilist can embrace Hasker's account of the power to perform a particular act under given circumstances.

And there are other perfectly reasonable accounts of such power open to the compatibilist. *Any* theorist will want to adopt certain constraints on an adequate account of the "particular" notion of power. The compatibilist, it seems to me, can adopt the constraint that, in ascribing such power to perform *A*, one holds fixed all temporally nonrelational facts about the past and all present circumstances that prevent (in Hasker's sense) the agent's performing *A*. Thus, the compatibilist (who claims that facts about God's beliefs are temporally relational) can embrace a "particular" notion of power. Thus, Hasker has not given us reason to think that the com-

---

claim that God's beliefs are soft facts must be given up, it would be overly optimistic to expect the compatibilist at this point to fold his tents and steal silently away. For compatibilists have also made claims about powers we have over the past—claims that need not depend on the distinction between hard and soft facts . . ." (95). This remark makes sense if Hasker is talking about hard facts in the sense of the first moment or aspect—temporal indifference (or better, being "over-and-done-with," i.e., being solely about the past). But, given the official approach of the book, I do not see how it makes sense at all: if facts about God's beliefs are hard (on the current definition), tent folding seems in order for the compatibilist.

[5] See Hasker's comments on 92–3, esp. fn. 19.

patibilist is committed to an implausible conception of power or that the compatibilist and incompatibilist are talking past each other.[6]

Hasker claims that the compatibilist cannot consistently affirm what he calls "libertarian free will." (By this he means the sort of free will that requires genuinely open alternative possibilities; this might be contrasted with an "actual-sequence" approach to free will and moral responsibility.) This claim allegedly follows from an account of libertarian free will conjoined with Hasker's claim that the compatibilist is committed to the general notion of power.

Here is Hasker's account of libertarian free will:

> (FW') N is free at *T* with respect to performing *A* = df. It is in N's power at *T* to perform *A* and it is possible at *T* for N to exercise that power, and it is in N's power at *T* to refrain from performing *A*, and it is also possible at *T* for N to exercise *that* power (138).

If one accepts (FW') and also the claim that the compatibilist is committed to the general notion of power, then one must conclude that the compatibilist cannot affirm libertarian free will. This is because an agent's having the general power is compatible with its not being possible that he exercise it and, indeed, if God holds the relevant belief it is (allegedly) not possible for the agent to exercise the relevant power.

Hasker says:

> It is clear . . . that a compatibilist . . is claiming that Peter can *have* the power to refrain from sinning *even though it is logically impossible that he should exercise that power under the existing circumstances*. But if one has the "power to do otherwise" only in that sense —the sense in which *having* the power does not guarantee that it is possible for the power to be *used*—then the central idea of libertarianism . . . has been lost (141).

But we saw above that the compatibilist need not embrace the general notion of power; rather, he can adopt a particularized notion. Also, the compatibilist can adopt a notion of possibility that permits the possibility of agents' exercising their powers, even if God exists. On the relevant notion of possibility, God's beliefs are not held fixed; but this is entirely appropriate (on the compatibilist's picture) insofar as God's beliefs are neither temporally nonrelational facts about the past nor present facts which are causally relevant to the agent's action.

---

[6] William P. Alston also believes that the compatibilist and incompatibilist are talking past each other; cf "Divine Foreknowledge and Alternative Conceptions of Human Freedom," *International Journal for Philosophy of Religion*, XVIII (1985): 19–32; repr. in Fischer, ed. Alston's view is, however, slightly different from Hasker's Alston does not attribute to the compatibilist the general notion of power; rather, he distinguishes between a compatibilist and incompatibilist "sense" of the particular notion of power.

Let us suppose that Peter sins at *T*. Is it possible for him to exercise his power to refrain from sinning? It is certainly impossible for him to exercise this power if *all* the "existing circumstances"—*including his sinning*—are held fixed. Thus, even an incompatibilist cannot hold such a stringent requirement. And when the requirement is relaxed (presumably employing the notion of soft facts), it seems that there is room for the compatibilist to assert that it is indeed possible for Peter to exercise this power. Thus, contrary to Hasker, there is no inconsistency in the compatibilist's affirming libertarian free will.

Wierenga systematically undertakes to give formulations of the various divine attributes, such as omnipotence, omniscience, perfect goodness, and so forth. He carefully sets out the accounts, and then evaluates them in the light of objections. In contrast to Hasker, Wierenga argues that God's omniscience is robust and includes foreknowledge of future free actions and middle knowledge.

In this review, I shall focus on Wierenga's discussion of the relationship between God's omniscience and human freedom. Whereas this discussion contains much that is of merit, I believe that it is vitiated by a problem similar to the one isolated above: Wierenga treats temporal nonrelationality (or being genuinely about the past) and fixity as different aspects of the generic notion of "accidental necessity." As with Hasker, this approach leads to certain confusions.

Here is a particularly glaring example of this sort of confusion. Wierenga criticizes a whole class of accounts of what he calls "accidental necessity," because they have as a consequence that certain facts such as "Socrates drank hemlock at *T*" turn out to be accidentally necessary. The problem, according to Wierenga, is that Alvin Plantinga has shown that there are possible circumstances in which such facts must not be considered accidentally necessary (99–100; 108).

I should point out, first, that the accounts in question (Freddoso, Widerker and Zemach, Fischer) all are putative accounts of temporal nonrelationality, not fixity *per se*. That is, they are all attempting to give accounts of hardness in the sense of temporal nonrelationality or being genuinely and solely about the relevant times. And facts such as "Socrates drank hemlock at *T*" are precisely this sort of fact. Now, whether all temporally nonrelational facts about the past are *fixed* is quite another matter, and Plantinga's examples pertain to this issue; thus, it is unfair to adduce these examples as part of a criticism of various accounts of temporal nonrelationality.

Further, let us consider Plantinga's examples more carefully. Plantinga[7] suggests that

> perhaps you will be confronted with a decision of great importance—so important that one of the alternatives is such that if you were to choose *it*, then the course of human history would have been quite different from what in fact it is. Furthermore, it is possible that if God had foreseen that you would choose *that* alternative, he would have acted very differently. Perhaps he would have created different persons; perhaps, indeed, he would not have created [Socrates]. So it is possible that there is an action such that it is within your power to perform it and such that if you were to perform it, then God would not have created [Socrates] (*ibid.*, p. 257).

Wierenga concludes that it is possible that a fact such as "Socrates drank hemlock at *T*" is not accidentally necessary (after *T*). Plantinga draws a similar conclusion. But even if we are considering the notion of "fixity" rather than temporal nonrelationality, this is dialectically unfair. For the dialectical context is one in which we are considering whether God's foreknowledge is compatible with human freedom. Thus, it is inappropriate to adduce in the manner of Plantinga and Wierenga an example in which it is alleged that some agent is free despite God's foreknowledge. Plantinga's examples, as far as I can see, make essential use of a foreknowing God (or at least some individual with relevantly similar attributes); but then they should not be invoked in a context in which it is controversial whether God's foreknowledge is compatible with human freedom. So, even if we restrict our focus to accidental necessity in the sense of fixity, the use of such examples is dialectically inappropriate.

Another worry. Wierenga points out that none of the accounts of accidental necessity he considers offers any reason to classify God's beliefs at a time as accidentally necessary. Thus, he concludes that the argument for incompatibilism based on the fixity of the past "must be judged to be a failure" (113). But I believe that Wierenga is too quick to leap to the compatibilistic conclusion. I believe that, even if God's beliefs at times are temporally relational facts about those times, they have a certain internal structure; further, this structure may include a temporally nonrelational or *hard* element. It may be that falsifying facts about God's beliefs requires affecting the hard internal component. Thus, there might be a perfectly good argument employing the fixity of the past to the conclusion that God's foreknowledge is incompatible with human freedom, even if God's beliefs are deemed soft facts.[8]

---

[7] Plantinga develops these examples in his "On Ockham's Way Out," *Faith and Philosophy*, III (1986) 235–69, repr. in Fischer, ed. Cf. Wierenga, 99.

Wierenga considers the response to the incompatibilist's argument that posits that God is atemporal. On this approach, His omniscience is not foreknowledge, and the argument from the fixity of the past cannot get off the ground. Wierenga argues against this atemporalist strategy that facts such as "God timelessly knows that Jones mows his lawn at *T*" obtain at times prior to *T* and can be considered hard facts about the times in question. (This point is also made by Plantinga (*op. cit.*, p. 239).)

But it seems to me that the facts in question are not hard facts about the relevant times. A hard fact about a time (in the sense of a temporally nonrelational fact about that time) is both *genuinely* and *solely* about that time. Some soft facts are genuinely but not solely about the times in question. Facts about God's timeless beliefs are not even genuinely about the relevant times: they do not allege that anything "really happens" at the relevant times. The distinction between hard and soft facts presupposes what Freddoso has called the "primacy of the pure present"; hard facts assert that something happens in the pure present, but facts about God's timeless beliefs make no such assertion. Thus, it is implausible to suppose that the Boethian strategy can be blocked in this fashion.

Finally, I remind the reader that, whereas Hasker argues that compatibilism (with a "particular" notion of "power") is incoherent, Wierenga defends it as true. For what it is worth, I disagree with them both: I believe that such compatibilism is coherent but false.

Both books are rigorous and contain many worthwhile and provocative arguments. They both apply sophisticated analytical tools to the subject matter. Despite the flaws noted above, I believe that they are enlightening and useful books. They constitute worthy additions to Cornell University Press's series, *Cornell Studies in the Philosophy of Religion.*

JOHN MARTIN FISCHER

University of California/Riverside

---

[8] I develop this sort of position in my "Hard-Type Soft Facts," *The Philosophical Review*, XCV (1986): 591–601.

Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. JUDEA PEARL. San Mateo: Morgan Kaufmann, 1988. 552 p. Cloth $49.95.

This is an extraordinary book. The author, an electrical engineer by trade, is writing for an audience of computer scientists concerned with artificial intelligence. Yet much of what he has to say is directly relevant to the concerns of philosophers. In fact, I would say that no philosopher concerned with probability, decision theory, defeasible logic, or even epistemology in its formal guises can afford to be ignorant of the contents of this book. Although it is written for researchers in artificial intelligence, and is formidably technical in some of its parts, it also contains much that is accessible to the nontechnical philosopher who possesses only a familiarity with elementary mathematics and logic.

The basic approach to uncertainty is a personalistic one. Probability is a function (a real-valued function) on statements. It is a subjective function, in the sense that there is no way in which an individual can "get it wrong." (Pearl does not discuss the problem of reconciling inconsistent personal probability assignments.) We can have the appearance of higher-order probabilities, because the proposition that an individual assigns a certain probability (has a certain belief in the occurrence of) an event is itself an event in the real world, and can have a probability.

This is not, nowadays, an unusual approach to probability, but Pearl seems to be quite unaware that this is hardly the "traditional" (358) approach. Similar jarring observations occur elsewhere. "Confidence intervals," for example, are used to refer to the intervals in which an individual's confidence lies, rather than to the common statistical construct of the same name. One should be aware of these things, but one should not complain: the adjustments to more conventional terminology are easy to make for those with the requisite background, which most philosophy students interested in these matters have.

The book covers a broad spectrum of issues, from a point of view, one component of which is the personalistic Bayesian view that a statement is probable because an agent thinks it so, and one component of which is to demand that only tractible computations are of serious interest. This latter point of view is an interesting one. Some decades ago what mattered in philosophy was what could be stated exactly and explicitly. That a certain sentence followed in a given logic from certain others was the paradigm: exact rules could be given. By the same token, that a certain sentence does *not* follow

0022-362X/91/8808/434–7

from certain others can be stated perfectly explicitly, even though there are no rules that tell us when this is the case.

We are being led, by artificial intelligence and computer science, to pay somewhat more attention to the processes by which we come to know that conditions such as "provability" are fulfilled. It is one thing to have a clear idea of what it is to be a sentential tautology. It is quite another matter, even though the question is perfectly decidable, to have a practical way of deciding whether an arbitrary formula, with, say, a few hundred components, is a tautology.

In the same way, it is one thing to adopt a subjectivistic or personalistic view of probability, and smilingly say: well, it is just a matter of computing your posterior; it is quite something else to explore ways of being able to do that in *reasonable* amounts of time. As a computer scientist, Pearl focuses on the latter. Whether philosophers should do likewise is an open question—one might most plausibly suppose that they ought to explore *all* questions—but it is certainly refreshing and important to be reminded of the difference between "in principle" and "in practice." An instance in which the distinction between in-principle definability and computational feasibility may be deeply important is the possible role of computational feasibility in determining how people actually think about the world. Pearl claims, for example, that "the human obsession with causation . . . is computationally motivated" (383). This idea is explored for several pages, and it certainly seems a fruitful one to consider, going far beyond the simple notion that people "invent" the causal relation. It may be able to explain how and when causal relations are interposed between observational events.

Pearl's book provides numerous techniques for getting on the right side of the line between tractable and intractable computation. One of the most important and effective is the technique of exploiting "conditional independence." This is a notion that should play a frequent and fundamental role in philosophy. Although the *idea* is familiar to philosophers—A and B are conditionally independent given C, if, once you know C, the fact that A occurred does not tell you anything new about B. For example, grain production and golf scores are not independent: high grain production tends to be associated with high golf scores; but if we know the weather, the association is explained away: given a rainy season, there will be a lot of grain produced, and there will be high golf scores, given a rainy season, quite independently of whether in addition there is a lot of grain produced.

It turns out that this idea of conditional independence plays an important role in many realms: defeasible reasoning, nonmonotonic logic, epistemic updating, knowledge representation, etc. It can also be exploited in developing decision trees. Many of these areas are explored in Pearl's rich book.

The first chapter of Pearl's book provides an introduction to the importance of probability in AI systems: something few philosophers will need to be persuaded of. It is not the existence of uncertainty that is important for philosophers, but the question of how to handle it, how it is related to classical axioms for probability. Chapters 2, 3, 4, and 5 concern the representation of belief and its computationally feasible revision. These chapters contain the technical meat of the book, as well as containing a serious look at issues that are not so comfortable for Pearl's approach. These chapters will perhaps be the most difficult going for philosophers. Part of the reason for this is the unfamiliarity of the terminology: some familiar philosophical ideas are given computer science names (278: ". . . more plausible, i.e. . . . has the highest posterior probability"). The same is true of the examples, for example, medical diagnosis, but here their unfamiliarity may provide the means for communicating new ideas.

Chapter 6, on decision and control, may well be one of the most rewarding in the book for philosophers. It is one thing to contemplate the abstract possibility of a complete decision tree for a moderately complex decision, but it is quite another matter to use these ideas in a responsible and tractable way to improve the quality of our decision making. There are no conceptual breakthroughs here—no challenges to the standard axioms. But the passage from abstract possibility to practical reality can have a profound effect on our view of decision theory. Here, as in other places, the one drawback of wordprocessing makes itself felt: inconsistency among the iterations of a manuscript. For example, a proof on page 296 refers to the "decomposability axiom," but there is no such axiom. One can figure out which axiom is intended among those a page or so earlier (axiom 3), but it is frustrating to have to do so. Such obstacles notwithstanding, the chapter is a brilliant one.

The seventh chapter will be seen as one of the weakest, from the point of view of the philosopher. Pearl has read widely and well in the philosophical literature as well as in the statistical literature. Nevertheless, there remains a certain incompleteness and inaccuracy that occasionally jars. The eighth ("Learning Structure from Data"), by contrast, provides material that is unfamiliar to most philosophers, and that should be familiar. It does not concern what we think of as induction, despite its title, but rather the problem of how best to structure a domain, given extensive statistical data about that domain. Take statistical inference for granted; suppose you have as much data as you want: now what can you learn about causal relations, about independence, and the like.

Chapter nine will give many philosophers a feeling of *deja vu.* Here again is Shafer's novel representation of uncertainty, this time dressed for artificial intelligence. As a thoroughgoing probabilist, Pearl must bend over backward to give the Shafer/Dempster formalism (as it is referred to in AI) its due. In fact, he bends over rather too far, and is more concerned to illustrate differences than to find commonalities. Thus he finds a "semantic" difference, in that Shafer's view is concerned with "degrees of provability," while probability is concerned with "degrees of belief." It is not clear why construing 'Bel' as the lower envelope of *certain* classes of probability functions does not resolve the differences; Pearl does not address this question.

The final chapter, on the relation between probability and various nonstandard logics: logics for defeasible reasoning, default logics, and the like, will strike many philosophers as the most fun and most informative. Many of the novelties of logic in artificial intelligence are revealed here; it is an area in which philosophers have much to contribute.

Finally, there is an extensive bibliography.

This is a great book. It is not always easy. Its difficulty is compounded by the fact that the writing and editing are not always smooth (a difficulty that may, by now, have been corrected in a new edition). It will nevertheless be a rewarding book for most philosophers, as well as statisticians, decision theorists, logicians, probabilists, and, of course, those in artificial intelligence.

HENRY E. KYBURG, JR

University of Rochester

Knowledge in Flux: Modeling the Dynamics of Epistemic States. PETER GÄRDENFORS. Cambridge: MIT Press, 1988. xi, 262 p. Cloth $27.50.

Recently an increasing number of writers have sought to direct the attention of epistemologists to the topic of belief change. Thus, Peter Gärdenfors declares his intention in the preface of his book to examine "how to change your mind." The result of his efforts is an outstanding piece of work.

0022-362X/91/8808/437–44

Discussing belief change presupposes a conception of the space of changes available to an inquirer in a given state of belief and, as a consequence, a characterization of the space of belief states to which the inquirer can move. Gärdenfors's discussion of this matter in chapter 2 offers an overview of diverse "models" that have been proposed by philosophers and computer scientists in recent years.

Most of Gärdenfors's discussion focuses on what I would call changes in "states of full belief." What is fully believed is taken for granted by the inquirer as settled, at least for the time being. There is "no doubt" that its negation is false (23). Hence, the inquirer is maximally certain that it is true and assigns it subjective or credal probability 1. Gärdenfors takes for granted, however, that full belief is not only sufficient for assigning probability 1 but necessary as well, and develops an overview of Bayesian models of epistemic states in chapter 2, section 7, and, on this assumption, accounts of change in probability judgment in chapter 5.[1]

Chapter 3 investigates some of the structural properties of changes in states of full belief due to adding new information (perhaps through observation, testimony, or inference) or removing information already taken for granted. In expansions, information is added to the state of full belief together with all of the implications thereof, regardless of whether this issues in inconsistency or not. Nothing is given up. Gärdenfors thinks inconsistency is "epistemic hell" (51), however. When an agent adds some information to his initial epistemic state which is inconsistent with it, the inquirer should not retain everything in the initial state as he would in expansion but should give up the least amount of information needed to avoid contradiction while adding the new items. Thus, when an inquirer changes his epistemic state due to the addition of new information, Gärdenfors calls it a "minimal revision." Minimal revisions are expansions when the new information added is consistent with the initial state; but, otherwise, minimal revisions require the inquirer to give up some old information in order to ensure consistency.

Sometimes the initial "input" demands that a proposition be given up without doing so in order to avoid the epistemic hell of inconsistency. Such changes are "contractions." Just as the addition of inconsistent information induces secondarily the removal of information, the deletion of information induces secondary deletion. Revision and contraction both involve ceasing to believe what one initially believed due to the pressure of adding information inconsistent with the initial doctrine, in the case of revision, or due to the pressure of initially removing information, in the case of contraction. Such secondary deletions call for choosing which portions of the old information to remove, and this involves evaluating assumptions one initially takes for granted (and, hence, initially assigns probability 1) with respect to their vulnerability to being given up or, alternatively, how well they are "entrenched" in one's initial state of belief. Gärdenfors recognizes that entrenchment is not a question of assessing how probable a claim is and, in chapter 4, he investigates some of the structural properties that evaluations of entrenchment should have. Chapter 5, as already stated, explores changes in probability judgment.

---

[1] I doubt the necessity of full belief for a judgment of probability 1 A person could reasonably assign probability 0 to the hypothesis that a piece of meat weighs exactly 3 pounds (and, hence, probability 1 to its negation) while remaining in doubt as to whether the weight is exactly 3 pounds or not. Indeed, the person could fully believe that the weight falls in the interval from 2.5 lbs to 3.5 pounds while being in doubt as to what the exact value is. He could do so while assigning 0 credal probability to every exact "point value" in the interval. All of this stands in contradiction to the necessity of full belief for probability 1.

Part II of the book, consisting of four further chapters, considers various applications of the ideas developed in the first five chapters. Chapter 6 develops the idea that propositional logic can be rationalized by imposing certain plausible requirements on how expansions ought to proceed. Chapter 7 examines belief-change approaches to the analysis of conditionals along lines pioneered by F. P. Ramsey. Chapter 8 contains an account of an epistemological approach to explanation where changes in belief, especially contractions, play a central role. Chapter 9 attempts a similar approach for causal beliefs.

The discussions are clear, informative, and responsible throughout, even when the positions advanced are controversial. The result is the best book-length overview of the formal aspects of belief change available.

Gärdenfors's favorite way of representing states of full belief linguistically associates with each potential state a set of sentences in the (regimented) language L. Gärdenfors supposes that L contains the standard truth-functional connectives, that it is governed by a consequence relation $\vdash$ between a set of sentences in L and a sentence in L. If A is a truth-functional tautology, then $\vdash A$. If $\vdash A \leftrightarrow B$ and $\vdash A$, then $\vdash B$. It is not the case that $\vdash \bot$. $\vdash$ satisfies the deduction theorem and compactness. He then defines a nonabsurd belief set to be a set of sentences closed under the consequence relation for L and consistent according to that consequence relation. He takes for granted that legitimate changes of states of full belief are representable as changes in nonabsurd belief sets.

Notice that changes in belief sets are understood as representations of changes in belief states. For Gärdenfors, the object of study is not transformations of linguistic entities but of "epistemic states" (states of full belief). In chapter 6, Gärdenfors exploits this idea to sketch a characterization of propositions that avoids explicating them either in terms of linguistic entities or possible worlds but rather construes them to be functions from epistemic states to epistemic states. Gärdenfors begins with a belief model taken to be an ordered pair ⟨K, Prop⟩ consisting of a set K of epistemic states and a set Prop of functions from K to K. Given an epistemic state K in K and a proposition A in Prop, $A(K)$ is intended to characterize the epistemic state $K_A{}^+$ which is the "expansion" of K by adding the proposition A. Gärdenfors introduces a system of assumptions and definitions about the kinds of functions in Prop which have this effect and which introduce the standard propositional connectives. In particular, given his assumptions, we can define a proposition B to be a consequence of a proposition A, if and only if for every K in K, $B(A(K)) = A(K)$. The postulates Gärdenfors introduces for the expansion functions in Prop suffice to induce a partial ordering of the epistemic states in K, although Gärdenfors does not explicitly say so. If B is a consequence of A, then $B(K)$ is a consequence of $A(K)$. Hence, one could have begun by characterizing a belief model as a set K of epistemic states and a consequence relation defined on this set. The expansions would then consist of all transitions from epistemic states to other such states that are consequences of the initial state. If, for some pair of epistemic states K and K', there is a greatest lower bound to the epistemic states that have both states as consequences (so that their "meet" exists), then the expansion of K by adding K' is that greatest lower bound or meet. Consequently, if we postulate that the meets of pairs of epistemic states always do exist, we can regard the epistemic states themselves as the propositions and represent the expansion of K by adding K' as the meet of these two epistemic states. Gärdenfors's approach does not guarantee that the meets of pairs of epistemic states always exist (though it does not preclude this). K is, as he explicitly states, completely unstructured except insofar as structure is induced by the functions in Prop. His postulates for Prop guarantee that the ranges of the functions in Prop (which are sets of epistemic states) have an algebraic structure. Given the seven postulates Gärdenfors finds compelling, they have the structure of a "pseudo Boolean algebra" sufficient to yield an "algebraic semantics" for intuitionistic propositional logic and, with an extra assumption, they have the structure of a Boolean algebra that yields classical propositional logic.

Gärdenfors's approach allows for the possibility that an inquirer may be in an epistemic state K and have available another state K' to which he might move without having available the conceptual opportunity of moving to the meet of these two states. In my judgment, this seeming generosity is, in point of fact, excessively restrictive on the conceptual opportunities available to the inquirer. One should not impose roadblocks in the path of inquiry by ruling out a shift to the meet. If this is right, there is no advantage in proceeding as Gärdenfors does, as compared to imposing the requisite algebraic structure (of a pseudo Boolean algebra or a Boolean algebra as the case might be) directly on the epistemic states and letting the epistemic states themselves serve as the propositions or "contents."

This reservation ought not to detract from Gärdenfors's achievement. If the topic is change in states of full belief, what we require in the way of logical structure ought to be derivable from our characterization of the range of epistemic states and the kinds of shifts that are thereby rendered available. Whether propositions are identified with epistemic states as I suggest or with expansion functions as Gärdenfors proposes, one can jettison much of the technical apparatus (such as the framework of possible worlds) that has dominated contemporary power without loss of power. Moreover, it becomes clear that linguistic representations of epistemic states by belief sets are just that—linguistic representations. Just as points in phase-space are used to represent mechanical states in classical mechanics without presupposing that mechanical systems speak the language of phase-space, the use of belief sets to represent epistemic states implies nothing about the way inquirers use language to manifest their epistemic states.

Gärdenfors resolutely eschews any concern with the concept of truth in his account of belief change. His pragmatism thereby resembles more closely the attitude of John Dewey than that of C. S. Peirce. This perhaps accounts, at least in part, for his ambivalence between resting content with an intuitionistic logic or a classical logic. For those like myself who see avoidance of error as a desideratum in inquiry, it is crucial that epistemic states or states of full belief be endowed with truth values, where truth or freedom of error is preserved under the consequence relation and where, for every epistemic state, there is a unique "classical negation" that is false (true), if and only if the first state is true (false). This kind of consideration argues in favor of requiring at a minimum that the epistemic states have the structure of a Boolean algebra and that the logic be classical. By allowing structures to be pseudo Boolean algebras that are not Boolean, Gärdenfors is recognizing that an agent may have a conceptual repertoire that is limited by the absence of negations for at least some epistemic states. For someone who thinks that avoidance of error is a desideratum in inquiry, the lack of an option of shifting to an epistemic state that is the classical negation of the initial one is an important deprivation of an epistemic option and, in Peirce's terms, does place roadblocks in the path of inquiry.

This objection presupposes that avoidance of error is an objective of inquiry. Given Gärdenfors's Deweyite indifference to truth, the worry being registered may not carry much weight with him.

Gärdenfors's belief sets (the linguistic representations of epistemic states) correspond to what elsewhere I have called "potential corpora" and modulo fine points of technical detail; there is very little difference between them. There is one important point on which we seem to disagree. Gärdenfors seeks an account of belief change that avoids countenancing any change to an absurd or inconsistent belief set or corpus. As a consequence, Gärdenfors has a rather different view from mine concerning what are the basic types of belief change.

I have advocated the view that all legitimate changes of states of full belief (changes of corpus) should be decomposed into sequences of legitimate expansions and contractions. It is a crucial feature of my view that sometimes expansion into inconsistency is legitimate. Gärdenfors's account of belief change seeks to preclude changes of this kind. He does so by taking revisions as basic. A revision of K by adding A is an expansion of K by adding A when A is consistent with K. It is what I have called a "replacement" in which $\sim A$ appears in the final state and A appears in the initial state otherwise. I deny that replacements are basic. To be legitimate, they must be decomposed into sequences of legitimate contractions and expansions. Gärdenfors does not, and, indeed, cannot require this.

The problem cases concern situations where the agent initially fully believes A and makes observations that yield reports that $\sim A$ is true or receives reliable expert testimony to this effect. According to my account, the agent expands into inconsistency from which he then seeks to extricate himself by contraction. He may do so by (i) questioning the process whereby he came to be sure that A is false, (ii) questioning the background assumption that it is true, or (iii) doing both. According to Gärdenfors's account, one never should expand into inconsistency in such cases. One should engage in a "minimal revision" that requires replacing A by its negation. This process will mimic my account when the contraction from inconsistency is of type (ii). Hence, if on my view the only type of legitimate contraction from inconsistency were of type (ii), the difference between Gärdenfors and myself would be a mere formal one. But I think it false to presystematic practice that, when one stumbles on a conflict between observation and background, one always questions the background rather than the observation. Nor do I think it unwise in some contexts to be skeptical of observation rather than or as well as the background. The interesting methodological issue is to identify the considerations that decide in favor of (i), (ii), or (iii) in any given context. Gärdenfors's approach precludes our addressing this issue by settling on a specific answer for all occasions in advance. For this reason, I think it an advantage to countenance expansions into inconsistency as sometimes legitimate and to regard expansion and contraction as basic.

Gärdenfors thinks of his investigations as addressing the "dynamics" of epistemic states. I think the analogy with physics is misleading in two important respects. First, as Gärdenfors clearly and rightly emphasizes, the epistemic states or states of full belief are best compared to equilibrium states. By representing such states by belief sets of sentences closed under a consequence relation, we are not assuming that inquirers are or prescribing that they ought to be logically omniscient. But as Gärdenfors rightly sees, the changes in view that occur when an agent recognizes a logical consequence of his current convictions, and becomes for this reason certain of its truth, are quite different in ways relevant to a systematic account of belief change from the changes that occur when an agent changes the set of convictions in a way that modifies their logical consequences. There are authors who do not recognize the difference as relevant or important. But Gärdenfors does, and I agree. Consequently, it is possible to regard the first kind of change as a change from disequilibrium to equilibrium and to envisage the inquirer as committed, while in disequilibrium, to realizing a certain equilibrium state. The second kind of change would then be regarded as a change in commitment to an equilibrium state. This is not the place to defend the superiority of this way of viewing things (which is Gärdenfors's way as well as mine) to views such as Gilbert Harman's. But if one embraces it, the account of belief change Gärdenfors and I are interested in is a study of changes from one equilibrium state to another. The principles regulating such changes (whether prescriptive or descriptive) are not analogs of dynamical principles but of principles of comparative statics. The name does not matter. But it is misleading to think of the kind of study Gärdenfors is engaged in, worthwhile as it is, as the analog of an investigation in dynamics.

This brings me to what, in my judgment, is the chief lacuna in Gärdenfors's magisterial essay. Gärdenfors restricts attention to changes where an initial "input" stipulates that a certain kind of change is to take place. The change may involve adding a proposition or removing one. In the case of expansion, addition of a proposition uniquely determines the new belief state. Just take the contents of the initial belief set and the sentence representing the new proposition and form the closure under the consequence relation. Gärdenfors bypasses the critical epistemological question of specifying conditions under which the agent is justified in adding some new proposition *A* to his initial belief state. This is not the usual foundationalist requirement that what is in the initial state be justified but only that the change contemplated be justified. According to the view of Dewey (whose views in some respects Gärdenfors seems to share), such changes are to be justified by looking at the goals of the inquirer and showing that the changes are best adapted to promoting the goals of the options available. There is no elaboration of this idea anywhere in Gärdenfors's book. The absence of attention to the decisional aspects of belief change carries over, at least in part, to Gärdenfors's discussion of contraction. Gärdenfors takes as input the demand to give up some proposition without considering the reasons that might justify attempts to gratify such demands. Gärdenfors recognizes the need to face the question of choosing between contraction strategies that remove the given proposition *A*, and has interesting things to say in this connection about entrenchment. This aspect of his view ought to fit naturally into a decision theoretic framework. Unfortunately, Gärdenfors does not recognize this. I conjecture that this failure accounts for his tendency to ignore difficulties with the so-called "recovery postulate" on contractions and with the "Harper identity" for revisions.

Even so, Gärdenfors has succeeded in making important contributions to understanding how to represent epistemic states and to identifying the salient properties of expansion, contraction, and revision. Space does not permit discussion of his views on conditionals and explanation; but his chapters on these matters ought to be taken seriously by philosophers of science and philosophers of mind. Overall, Gärdenfors's essay is a major contribution of interest to epistemologists, philosophers of science, philosophers of mind, students of artificial intelligence, and psychologists. It ought to change our minds about how to change our minds.

ISAAC LEVI

Columbia University

NEW BOOKS: ANTHOLOGIES

ABOULAFIA, MITCHELL, ed.: *Philosophy, Social Theory, and the Thought of George Herbert Mead.* Albany: SUNY, 1991. xx, 319 p. Cloth $39.50, paper $12.95.

ADAMS, MARILYN MCCORD, and ROBERT MERRIHEW ADAMS, eds.: *The Problem of Evil.* New York: Oxford, 1990. viii, 230 p. Cloth $59.00, paper $15.95.

AL-HIBRI, AZIZAH Y., and MARGARET A. SIMONS, eds.: *Hypatia Reborn: Essays in Feminist Philosophy.* Bloomington: Indiana UP, 1990. xii, 349 p. Cloth $39.95, paper $18.95.

ANNAS, JULIA, ed.: *Oxford Studies in Ancient Philosophy: Volume VIII 1990.* New York: Oxford, 1990. ix, 301 p. Cloth $34.50.

BEDAU, HUGO ADAM, ed.: *Civil Disobedience in Focus.* New York: Routledge, 1991. vi, 217 p. Cloth $69.95, paper $17.95.

BENHABIB, SEYLA, and FRED DALLMAYR, eds.: *The Communicative Ethics Controversy.* Cambridge: MIT, 1990. viii, 378 p. Cloth $37.50, paper $16.95.

BENTHAM, JEREMY: *Securities Against Misrule and other Constitutional Writings for Tripoli and Greece.* PHILIP SCHOFIELD, ed. New York: Oxford, 1990. li, 326 p. Cloth $98.00.

BERNASCONI, ROBERT, and SIMON CRITCHLEY, eds.: *Re-Reading Levinas.* Bloomington: Indiana UP, 1991. xviii, 252 p. Cloth $39.95, paper $14.95.

CARD, CLAUDIA, ed.: *Feminist Ethics.* Lawrence: Kansas UP, 1991. viii, 300 p. Cloth $29.95, paper $14.95.

COVER, J. A., and MARK KULSTAD, eds.: *Central Themes in Early Modern Philosophy: Essays Presented to Jonathan Bennett.* Indianapolis: Hackett, 1990. xiii, 336 p. Cloth $32.50, paper $16.95.

DAHLSTROM, DANIEL O., ed.: *Philosophy and Art.* Washington: Catholic UP, 1991. xii, 266 p. Cloth $44.95.

FISCH, MENACHEM, and SIMON SCHAFFER, eds.: *William Whewell: A Composite Portrait.* New York: Oxford, 1991. xiv, 403 p. Cloth $98.00.

FLANAGAN, OWEN, and AMELIE OKSENBERG RORTY, eds.: *Identity, Character, and Morality: Essays in Moral Psychology.* Cambridge: MIT, 1990. vii, 487 p. Cloth $37.50.

GEORGE, ALEXANDER, ed.: *Reflections on Chomsky.* Cambridge: Blackwell, 1989. xii, 271 p. Cloth $51.95, paper $19.95.

GOICOECHEA, DAVID, JOHN LUIK, and TIM MADIGAN, eds.: *The Question of Humanism: Challenges and Possibilities.* Buffalo: Prometheus, 1991. 341 p. Cloth $29 95.

HUNT, G. M. K., ed.: *Philosophy and Politics.* New York: Cambridge, 1990. vi, 147 p. Paper $16.95.

JOHNSON, GALEN A., and MICHAEL B. SMITH, eds.: *Ontology and Alterity in Merleau-Ponty.* Evanston: Northwestern UP, 1990. xxxiv, 206 p. Cloth $39.95, paper $12.95.

KEYT, DAVID, and FRED D. MILLER, eds.: *A Companion to Aristotle's* Politics. Cambridge: Blackwell, 1991. xii, 407 p. Cloth $63.95., paper $26.95.

LAMBERT, KAREL, ed.: *Philosophical Applications of Free Logic.* New York: Oxford, 1991. x, 309 p. Cloth $49.95.

LEPORE, ERNEST, and ROBERT VAN GULICK, eds.: *John Searle and His Critics.* Cambridge: Blackwell, 1991. xiv, 405 p. Cloth $44.95.

MACLEAN, IAN, ALAN MONTEFIORE, AND PETER WINCH, eds.: *The Political Responsibility of Intellectuals.* New York: Cambridge, 1991. xviii, 311 p. Cloth $49.50, paper $15.95.

MACNIVEN, DON, ed.: *Moral Expertise: Studies in Practical and Professional Ethics.* New York: Routledge, 1990. xix, 231 p. Cloth $55.00.

MORTLEY, RAOUL, ed.: *French Philosophers in Conversation.* New York: Routledge, 1991. 108 p. Paper $12.95.

PAKALUK, MICHAEL, ed.: *Other Selves: Philosophers on Friendship.* Indianapolis: Hackett, 1991. xiv, 273 p. Cloth $29.50, paper $7.95.

PERKINS, ROBERT L., ed.: *The Corsair Affair: International Kierkegaard Commentary: Volume 13.* Macon, GA: Mercer UP, 1990. xxv, 193 p. Cloth $24.95.

PETERSON, GRETHE B., ed.: *The Tanner Lectures on Human Values, vol. XII.* Salt Lake City: Utah UP, 1991. vi, 390 p. Cloth $30.00.

PINSENT, DAVID HUME: *A Portrait of Wittgenstein as a Young Man: From the Diary of David Hume Pinsent 1912–1914.* G. H. VON WRIGHT, ed Cambridge: Blackwell, 1990. xxi, 119 p. Cloth $29.95.

REISS, HANS, ed.: *Kant: Political Writings.* New York: Cambridge, 1991. xv, 309 p. Cloth $44.50, paper $14.95.

ROSENTHAL, DAVID, ed.: *The Nature of Mind.* New York: Oxford, 1991. x, 642 p. Cloth $49.95, paper $19.95.

ROUSSEAU, G. S., ed.: *The Languages of Psyche: Mind and Body in Enlightenment Thought.* Berkeley: California UP, 1991. xix, 480 p. Cloth $55.00, paper $16.95.

SAYERS, SEAN, and PETER OSBORNE, eds.: *Socialism, Feminism and Philosophy: A Radical Philosophy Reader.* New York: Routledge, 1990. ix, 306 p. Cloth $69.95, paper $16.95.

SILVERMAN, HUGH J., ed.: *Writing the Politics of Difference.* Albany: SUNY, 1991. xvi, 372 p. Cloth $59.50, paper $19.95.

SIMON, YVES R.: *Practical Knowledge.* Robert J. Mulvaney, ed. New York: Fordham, 1991. xiii, 163 p. Cloth $35.00, Paper $19.95.

VILLANUEVA, ENRIQUE, ed.: *Information, Semantics, and Epistemology.* Cambridge: Blackwell, 1990. viii, 246 p. Cloth $49.95.

NEW BOOKS: PAPERBACKS

BATAILLE, GEORGES: *The Accursed Share: Volume 1: Consumption.* ROBERT HURLEY, trans. New York: Zone, 1991. 197 p. $12.95.

BRICKHOUSE, THOMAS C., and NICHOLAS D. SMITH: *Socrates on Trial.* Princeton: University Press, 1990. xiv, 337 p. $14.95.

BRAUDE, STEPHEN E.: *The Limits of Influence: Psychokinesis and the Philosophy of Science.* New York: Routledge, 1991. xiv, 311 p. $15.95.

CARRUTHERS, PETER: *Introducing Persons: Theories and Arguments in the Philosophy of Mind.* London: Routledge, 1991. viii, 265 p. $16.95.

CRANSTON, MAURICE: *Jean-Jacques: The Early Life and Work of Jean-Jacques Rousseau 1712–1754.* Chicago: University Press, 1991. 382 p. $17.95.

GELLNER, ERNEST: *Plough, Sword, and Book: The Structure of Human History.* Chicago: University Press, 1990. 288 p. $14.95.

HEIDEGGER, MARTIN: *Nietzsche: Volumes I and II.* DAVID FARRELL KRELL, trans. and ed. *Nietzsche: Volumes III and IV.* FRANK A. CAPUZZI, DAVID FARRELL KRELL, and JOAN STAMBAUGH, trans., DAVID FARRELL KRELL, ed. San Francisco: Harper Collins, 1991. 589 p. $16.95 (each).

HICK, JOHN: *An Interpretation of Religion: Human Responses to the Transcendent.* New Haven: Yale, 1991. xv, 412 p. $18.95.

JEFFREY, RICHARD C.: *The Logic of Decision:* 2nd ed. Chicago: University Press, 1990. xiv, 231 p. $14.95.

KIERKEGAARD, SØREN: *For Self-Examination: Judge for Yourself!* HOWARD V. HONG and EDNA H. HONG, trans. and ed. Princeton: University Press, 1991. xviii, 296 p. $14.95.

KURTZ, PAUL: *The Transcendental Temptation: A Critique of Religion and the Paranormal.* Buffalo: Prometheus, 1991. xvi, 500 p. $16.95.

MACGREGOR, GEDDES: *Dictionary of Religion and Philosophy.* New York: Paragon, 1991. xi, 696 p. $19.95.

PRICE, A. W.: *Love and Friendship in Plato and Aristotle.* New York: Oxford, 1990. xiv, 264 p. $18.95.

RAPAPORT, HERMAN: *Heidegger and Derrida: Reflections on Time and Language.* Lincoln: Nebraska UP, 1991. ix, 293 p. $10.95.

RAZ, JOSEPH: *Practical Reason and Norms.* Princeton: University Press, 1990. 220 p. $12.95.

RORTY, AMELIE OKSENBERG: *Mind in Action: Essays in the Philosophy of Mind.* Boston: Beacon, 1988. x, 378 p. $16.95.

ROSEN, STANLEY: *The Ancients and the Moderns: Rethinking Modernity.* New Haven: Yale, 1989. x, 236 p. $12.95.

SHANKER, S. G., ed.: *Godel's Theorem in Focus.* New York: Routledge, 1990. ix, 261 p. $14.95.

NEW BOOKS: TRANSLATIONS

ADORNO, THEODOR W.: *Notes to Literature: Volume One.* SHIERRY WEBER NICHOLSEN, trans. New York: Columbia, 1991. xv, 284 p. Cloth $35.00.

BAKUNIN, MICHAEL: *Statism and Anarchy.* MARSHALL SHATZ, trans and ed. New York: Cambridge, 1990. xlix, 243 p. Cloth $44.50, paper $14.95.

BLOCH, ERNST: *Heritage of Our Times.* NEVILLE and STEPHEN PLAICE, trans. Berkeley: California UP, 1991. xiv, 377 p. Cloth $40.00.

BOURDIEU, PIERRE: *The Political Ontology of Martin Heidegger.* PETER

COLLIER, trans. Stanford: University Press, 1991. viii, 138 p. Cloth $22.50.

BOUSSUET, JACQUES-BENIGNE: *Politics Drawn from the Very Words of Holy Scripture.* PATRICK RILEY, trans. and ed. New York: Cambridge, 1990. lxxix, 415 p. Cloth $59.50.

CICERO. *On Duties.* M. T. GRIFFIN and E. M. ATKINS, eds. and trans. New York: Cambridge, 1991. li, 189 p. Cloth $37.50, paper $9.95.

DAVOUST, EMMANUEL: *The Cosmic Water Hole.* BARBARA JACHOWICZ, trans. Cambridge: MIT, 1991. xv, 206 p. Cloth $19.95.

DELEUZE, GILLES: *Empiricism and Subjectivity: An Essay on Hume's Theory of Human Nature.* CONSTANTIN V. BOUNDAS, trans. New York: Columbia, 1991. xi, 163 p. Cloth $32.50.

DUVE, THIERRY DE: *Pictorial Nominalism: On Marcel Duchamp's Passage from Painting to the Readymade.* DANA POLAN, trans. Minneapolis: Minnesota UP, 1991. xxii, 222 p. Cloth $39.95, paper $15.95.

HONNETH, AXEL, and HANS JOAS, eds.: *Communicative Action: Essays on Jürgen Habermas's* The Theory of Communicative Action. JEREMY GAINES and DORIS L. JONES, trans. Cambridge: MIT, 1991. ix, 301 p. Cloth $39.95.

KELSEN, HANS: *General Theory of Norms.* MICHAEL HARTNEY, trans. New York: Oxford, 1991. lx, 465 p. Cloth $110.00.

JACQUES, FRANCIS: *Difference and Subjectivity: Dialogue and Personal Identity.* ANDREW ROTHWELL, trans. New Haven: Yale, 1991. xxix, 374 p. Cloth $40.00.

JOHN OF SALISBURY: *Policraticus: Of the Frivolities of Courtiers and the Footprints of Philosophers.* CARY J. NEDERMAN, trans. and ed. New York: Cambridge, 1990. xxix, 240 p. Cloth $44.50, paper $14.95.

KLOSSOWSKI, PIERRE: *Sade My Neighbor.* ALPHONSO LINGIS, trans. Evanston: Northwestern UP, 1991. xv, 144 p. Cloth $24.95, paper $9.95.

KOFMAN, SARAH: *Freud and Fiction.* SARAH WYKES, trans. Boston: Northeastern UP, 1991. ix, 196 p. Cloth $35.00.

KRISTEVA, JULIA: *Strangers to Ourselves.* LEON S. ROUDIEZ, trans. New York· Columbia, 1991. xi, 230 p. Cloth $29.00.

PHILOPONUS and SIMPLICIUS: *Place, Void, and Eternity: Philoponus: Corollaries on Place and Void: Simplicius: Against Philoponus on the Eternity of the World.* DAVID FURLEY and CHRISTIAN WILDBERG, trans. Ithaca: Cornell, 1991. 153 p. Cloth $44.50.

SARTRE, JEAN-PAUL: *Critique of Dialectical Reason: Volume Two.* QUENTIN HOARE, trans. New York: Verso, 1991. x, 467 p. Cloth $60.00.

URVOY, DOMINIQUE: *Ibn Rushd (Averroes).* OLIVIA STEWART, trans. New York: Routledge, 1991. iv, 156 p. Paper $18.95.

Philosophical Inquiries

What Can She Know?

Feminist Theory and the Construction of Knowledge

By LORRAINE CODE. "A splendid book. Code surveys and discusses a great range of recent feminist writings. She also has a distinctive line of her own. This is feminist philosophy at its very best." —Annette C. Baier, University of Pittsburgh. $42.50 cloth, $14.95 paper

Unreality

The Metaphysics of Fictional Objects

By CHARLES CRITTENDEN. Drawing on the tools of Wittgensteinian philosophy and speech act theory, Crittenden argues that we can and often do make reference to *unreal* objects such as fictional characters, though they do not exist. $26.50 cloth, $9.95 paper

Whose Science? Whose Knowledge?

Thinking from Women's Lives

By SANDRA HARDING. "An exciting new contribution. It expands and refines Harding's earlier arguments concerning gender, and breaks new ground in its consideration of race and science. It is must reading."—Anne Fausto-Sterling, author of *Myths of Gender*. $34.50 cloth, $12.95 paper

Time and Eternity

By BRIAN LEFTOW. Leftow makes an important contribution to the longstanding debate among philosophers and theologians about the nature of God's eternity. The author argues that God is eternal in that he exists timelessly; that is, that though God exists, he does not exist at any time. *Cornell Studies in the Philosophy of Religion*. $42.95

Thomas Reid on Freedom and Morality

By WILLIAM L. ROWE. In this succinct and well-written book, one of our most eminent philosophers provides a fresh reading of the view of freedom and morality developed by Thomas Reid (1710–1796). $26.95

John Dewey and American Democracy

By ROBERT B. WESTBROOK. "Without a doubt the finest, most comprehensive and informative book on Dewey that has ever been written." —Richard J. Bernstein, New School for Social Research. Westbrook here reconstructs the evolution of John Dewey's thought and practice, combining readings of his major works with engaging accounts of his activism. $29.95

Place, Void, and Eternity

By PHILOPONUS and SIMPLICIUS. Translated by DAVID FURLEY and CHRISTIAN WILDBERG. This volume makes available in translation two key documents from the commentaries on Aristotle's *Physics*: Philoponus' *Corollaries on Place and Void* and a polemical excerpt in which Simplicius seeks to refute Philoponus' case against Aristotle's concept of an eternal world. *Ancient Commentators on Aristotle*. $44.50

Cornell University Press

124 Roberts Place, Ithaca, NY 14850

Oxford

Care and Commitment
Taking the Personal Point of View

JEFFREY BLUSTEIN, *Mercy College*

In this book, Blustein presents the first study of an ethics of care, offering a detailed exploration of human "care" in its various guises: concern for and commitment to individuals, ideals, and causes. Blustein focuses on the nature and value of personal integrity and intimacy, and on the questions they raise for traditional moral theory.

1991 288 pp $32 50

Experts in Uncertainty
Expert Opinion and Subjective Probability in Science

ROGER M. COOKE, *Delft University of Technology, The Netherlands*

An extensive survey and critical examination of the literature on the use of expert opinion in scientific inquiry and policy making, this book considers how expert opinion is being used today, how an expert's uncertainty is or should be represented, how people do or should reason with uncertainty, how the quality and usefulness of expert opinion can be assessed, and how the views of several experts might be combined.

1991 448 pp $65 00

Acceptable Evidence
Science and Values in Risk Management

Edited by DEBORAH G. MAYO, *Virginia Polytechnic Institute and State University*, and RACHELLE HOLLANDER, *National Science Foundation, Washington, D C.*

Concentrating on the entry of values in collecting, interpreting, communicating, and evaluating the evidence of risks, this volume analyzes environmental and medical controversies, and assumptions underlying views about risk assessment and the scientific and statistical models used in risk management.

1991 288 pp $35.00

Character

JOEL J. KUPPERMAN, *University of Connecticut, Storrs*

What is character? How does it relate to having a self, or to the process of moral decision? Are we responsible for our characters? *Character* answers these questions, and goes on to examine the place of character in ethical philosophy. Kupperman demonstrates the deficiencies of a number of classic and contemporary ethical theories that do not take account of the idea of character, and offers his own character-based theory.

1991 240 pp $29 95

Existential Epistemology
A Heideggerian Critique of the Cartesian Project

JOHN RICHARDSON, *New York University*

"Particularly noteworthy since it shows why any epistemology that fails to understand its own existential roots is bound to end up in skepticism....It is a welcome addition to a small body of work that attempts to bring together basic currents in the Anglo-American and European philosophical traditions."—*Choice*

Paper 1991 224 pp paper $18 95 cloth $49 95

The Philosophy of Vacuum

Edited by SIMON W. SAUNDERS, *Harvard University*, and HARVEY R. BROWN, *Oxford University*

This collection brings together philosophically-minded specialists who engage the issues surrounding the concept of vacuum in the context of classical gravity, quantum electrodynamics, and the grand unification program. A work in modern metaphysics, in which the concepts of substance and space interweave in the most intangible of forms, this book provides the background and context of our physical experience: vacuum, void, or nothingness.

1991 304 pp $39.00

Prices are subject to change and apply only in the U.S.
To order, send check or money order to: Humanities Marketing, Dept. MG
To order by phone using major credit cards please call
(212) 679-7300, ext. 7106

Oxford University Press
200 Madison Avenue • New York, NY 10016

Centenary Volume!

MIND

A quarterly review of philosophy

Editor: Professor Mark Sainsbury, King's College, London

MIND is the best known British philosophical journal, and one of the most widely read and respected journals in professional philosophy. Founded in 1976, for over a century it has represented the leading philosophical ideas of its time. In 1991 MIND celebrates the 100th Volume of the New Series, and is publishing a special centenary issue to mark the occasion.

Centenary Issue, October 1991

This special book-length issue (approx 350 pages) will be devoted to the subject of Mind and Content.

Contributors will include:

• Christopher Peacocke • Ruth Millikan • Alan Millar • Colin McGinn • James Hopkins • • Andrew Woodfield • Lord Quinton • • Philip Pettit • Martin Davies •

This issue is included in the price of the 1991 subscription, or available as a single issue

Subscription rates Volume 100, 1991

(4 issues including centenary issue)

Individuals: UK & Europe £15.00; Rest of World US$30.00

Institutions: UK & Europe £25.00; Rest of World US$47.00

To subscribe, obtain further information or a free sample copy of MIND, please write to: Journals Marketing Department (PJ), **Oxford University Press, Inc.**,
2001 Evans Road, Cary, NC 27513

OXFORD UNIVERSITY PRESS

Cambridge University Press

Cambridge Texts in the History of Political Thought

Raymond Geuss, Quentin Skinner, and Richard Tuck, Series Editors

This series makes available the most important texts required for an understanding of the history of political thought. The principle aim is to produce new versions of the major texts themselves, the preference always being for complete texts. A special feature of the series is to complement individual texts, within the compass of a single volume, with subsidiary contextual material. Each volume contains an introduction on the historical identity and contemporary significance of the text concerned.

New Books in the Series:
Cambridge Texts in the History of Political Thought

Kant: Political Writings
Second Edition
H.S. Reiss, Editor
H.B. Nisbet, Translator

Two important texts illustrating Kant's view of history; his reviews of Herder's *Ideas on the Philosophy of the History of Mankind* and *Conjectures on the Beginning of Human History*; and the essay *What is Orientation in Thinking?* are added to this newly expanded edition.

39185-7 Hardcover $44.50
39837-1 Paper $14.95

Hobbes: *Leviathan*
Richard Tuck, Editor

A masterly analysis of how and why power must be used in a market society to preserve peace and prosperity.

39492-9 Hardcover $49.50
39641-7 Paper $10.95

Milton: Political Writings
Martin Dzelzainis, Editor
Claire Gruzelier, Translator

The greatest English Renaissance poet also devoted twenty years to prose writing in the advancement of religious, civil and political liberties.

34394-1 Hardcover $44.50
34866-8 Paper $14.95

Filmer: *Patriarcha* and Other Writings
Johann P. Sommerville, Editor

This new edition includes the principal writings of Sir Robert Filmer (1588–1653) as well as a chronology, bibliography and notes.

37491-X Hardcover $44.50
39903-3 Paper $14.95

Cicero: *On Duties*
Miriam Griffin and Margaret Atkins, Editors

De Officiis, Cicero's last theoretical work on the political and ethical values of the Roman governing class in the late Republic, is made more meaningful through an explanation of its relationship to its own time and place.

34338-0 Hardcover $37.50
34835-8 Paper $9.95

Bakunin: *Statism and Anarchy*
Marshall Shatz, Editor

The Russian anarchist's last work, written in the 1873 aftermath of the rise of the German Empire and the clash with Karl Marx in the first International.

36182-6 Hardcover $44.50
36973-8 Paper $14.95

Cambridge University Press

John of Salisbury: *Policraticus*
Of the Frivolities of Courtiers and the Footprints of Philosophers
Cary J. Nederman, Editor
Regarded as the first complete work of political theory of the Latin Middle Ages; for the first time in translation.
36399-3 Hardcover $44.50
36701-8 Paper $14.95

Bossuet: *Politics Drawn From the Very Words of Holy Scripture*
Patrick Riley, Editor
This is the first ever English rendition of the classic statement of divine right absolutism in French, published in 1707.
36237-7 Hardcover $59.50

Also available...

Bentham: *A Fragment on Government*
J.H. Burns and H.L.A. Hart, Editors
35064-9 Hardcover $37.50
35929-5 Paper $7.95

Montesquieu: *The Spirit of the Laws*
Anne M. Cohler, et al., Editors
36183-4 Hardcover $59.50
36974-6 Paper $16.95

J.S. Mill: *On Liberty:* with The Subjection of Women and Chapters on Socialism
Stefan Collini, Editor
37015-9 Hardcover $37.50
37917-2 Paper $7.95

Aristotle: *The Politics*
Stephen Everson, Editor
35449-8 Hardcover $37.50
35731-4 Paper $7.95

Constant: Political Writings
Biancamaria Fontana, Editor
30836-2 Hardcover $44.50
31632-4 Paper $14.95

Paine: Political Writings
Bruce Kuklick, Editor
36665-8 Hardcover $32.95
36678-X Paper $7.95

Locke: *Two Treatises of Government*
Peter Laslett, Editor
35448-X Hardcover $44.50
35730-6 Paper $8.95

More: *Utopia*
George M. Logan and Robert M. Adams, Editors
34573-1 Hardcover $39.50
34797-1 Paper $7.95

Richard Hooker: *Of The Laws of Ecclesiastical Polity:* Preface, Book I and Book VIII
A.S. McGrade, Editor
37011-6 Hardcover $44.50
37908-3 Paper $14.95

Leibniz: Political Writings
Patrick Riley, Editor
35380-7 Hardcover $44.50
35899-X Paper $14.95

Machiavelli: *The Prince*
Quentin Skinner and Russell Price, Editors
34240-6 Hardcover $39.50
34993-1 Paper $7.95

Available in bookstores or write:
CAMBRIDGE UNIVERSITY PRESS
40 West 20th Street, New York, NY 10011. Call toll-free 800-872-7423
MasterCard/VISA accepted. Prices subject to change.

HARVARD

A Progress of Sentiments

Reflections on Hume's *Treatise*

Annette Baier

In this work, Annette Baier argues that Hume's *Treatise of Human Nature* is a carefully crafted literary and philosophical work which itself displays a philosophical progress of sentiments. She describes how, by turning philosophy toward human nature, Hume initiated a new philosophy, a broader discipline of reflection that can embrace Darwin and Foucault as well as Freud.

$39.95 cloth

Studies in the Way of Words

Paul Grice

"[Grice's] work culminated in the William James lectures delivered at Harvard in 1967, and philosophers will be greatful for having them finally available in one volume together with many other of Grice's papers, and a retrospective epilogue." —Simon Blackburn, *Times Literary Supplement*

$14.95 paper

Contingencies of Value

Alternative Perspectives for Critical Theory

Barbara Herrnstein Smith

"Smith's book argues, very lucidly and persuasively, for a Deweyan conception of value. [It] is a very useful contribution to the philosophical literature on the topic." —Richard Rorty

$12.95 paper

The Logical Basis of Metaphysics

Michael Dummett

"This book is important, daring, controversial, and very deep. I believe that [Dummett's conclusions] mark one of the true high-water marks of twentieth-century philosophy."

—Hilary Putnam

William James Lectures

$34.95 cloth

Varieties of Moral Personality

Owen Flanagan

"Flanagan's book comes as a welcome link between previously isolated enterprises. It skillfully bridges philosophy, psychology and religion by the happy strategy of focusing on exemplary moral agents."

—Harvey Cox, Harvard Divinity School

$34.95 cloth

RECENTLY PUBLISHED

The Realm of Rights

Judith Jarvis Thomson

$35.00 cloth

Realism with a Human Face

Hilary Putnam

Edited and Introduced by James Conant

$30.00 cloth

The Concept of Logical Consequence

John Etchemendy

$25.00 cloth

Harvard University Press

Cambridge, MA 02138

(617) 495-2480

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NUMBER 9
SEPTEMBER 1991

page
449 *Empirical Equivalence and Underdetermination* Larry Laudan and Jarrett Leplin
473 *Penny Pinching and Backward Induction* Martin Hollis
BOOK REVIEWS
489 T. H. Irwin: *Aristotle's First Principles* Gisela Striker
497 Rüdiger Bittner: *What Reason Demands* Thomas E. Hill, Jr.
501 NOTES AND NEWS
502 NEW BOOKS

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1991 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027.

THE JOURNAL OF PHILOSOPHY
VOLUME LXXXVIII, NO 9, SEPTEMBER 1991

EMPIRICAL EQUIVALENCE AND UNDERDETERMINATION*

During this century, there emerged from the philosophical analysis of scientific theories two results invested with broad epistemological significance. By the 1920s, it was widely supposed that a perfectly general proof was available for the thesis that there are always empirically equivalent rivals to any successful theory. Secondly, by the 1940s and 1950s, it was thought that—in large part because of empirical equivalence—theory choice was radically underdetermined by any conceivable evidence. Whole theories of knowledge (e.g., W. V. Quine's[1]) have been constructed on the presumption that these results were sound; at the same time, fashionable recent repudiations of the epistemic project (e.g., Richard Rorty's) have been based on the assumption that these results are not only legitimate, but laden with broad implications for the theory of knowledge.

In this paper, we reject both the supposition of empirical equivalence and the inference from it to underdetermination. Not only is there no general guarantee of the possibility of empirically equiva-

---

* The authors wish to thank Bas van Fraassen, Alan Goldman, William Lycan, Jim Maffie, David Stump, and Mary Tiles for criticism.

[1] The holistic theses of Pierre Duhem and Quine are probably the major instigators of this line of thought, as it appears in contemporary epistemology and philosophy of science. Quine's doctrines of the inscrutability of reference and the indeterminacy of translation are forms of underdetermination that depend on holism. See, e.g., "Epistemology Naturalized," *Ontological Relativity and Other Essays* (New York: Columbia, 1969), pp. 80ff.

The antirealist arguments of Arthur Fine and van Fraassen depend, in turn, on the thesis of underdetermination. See van Fraassen, *The Scientific Image* (New York: Oxford, 1980), ch. 3; and Fine, "Unnatural Attitudes: Realist and Instrumentalist Attachments to Science," *Mind*, xcv (April 1986): 149–79.

In the nineteenth century, both J. S. Mill and W. Whewell treated the possibility of empirical equivalence as an obstacle to scientific knowledge, differing as to whether and how it could be overcome.

0022-362X/91/8809/449–72

lent rivals to a given theory, but empirical equivalence itself is a problematic notion without safe application. Moreover, the empirical equivalence of a group of rival theories, should it obtain, would not by itself establish that they are underdetermined by the evidence. One of a number of empirically equivalent theories may be uniquely preferable on evidentially probative grounds. Having argued for these conclusions in the first two sections, respectively, we shall propose, in section III, a diagnosis of the difficulty that has impeded their recognition, and extract an attendant, positive moral for the prospects of epistemology.

I PROBLEMS WITH EMPIRICAL EQUIVALENCE

A. Inducements to skepticism. The idea that theories can be empirically equivalent, that in fact there are indefinitely many equivalent alternatives to any theory, has wreaked havoc throughout twentieth-century philosophy. It motivates many forms or relativism, both ontological and epistemological, by supplying apparently irremediable pluralisms of belief and practice. It animates epistemic skepticism by apparently underwriting the thesis of underdetermination. In general, the supposed ability to supply an empirically equivalent rival to any theory, however well supported or tested, has been assumed sufficient to undermine our confidence in that theory and to reduce our preference for it to a status epistemically weaker than warranted assent.

Specifically, this supposed ability is the cornerstone of arguments for the inscrutability of reference and the indeterminacy of translation, which together insulate the epistemic agent by challenging the objectivity of criticism on which an entire philosophical culture has depended. It has spawned prominent, contemporary versions of empiricism, including those of Quine, Bas van Fraassen, and J. D. Sneed, which belie the promise of science to deliver theoretical knowledge. It encourages conventionalism in geometry through Hans Reichenbach's invocation of universal forces. It questions the possibility of ordinary knowledge of other minds through the contrivance of the inverted spectrum. It blocks inductive generalization through the stratagem of fashioning artificial universals to vie with natural kinds, as in Nelson Goodman's "grue" paradox, reducing the status of apparent laws to mere entrenchment.

The linkage between empirical equivalence and epistemic skepticism has roots that go back well beyond such contemporary manifestations. Hume reduced causal judgments to psychological projections of habit by offering coincidental concomitance as the empirically equivalent alternative to natural necessity. Descartes worried that, were our impressions illusory or demonically fabricated, what we take to be evidence would come out the same, inferring that empirical beliefs are unwarranted while those possibilities are open. Berkeley similarly exploited empirical equivalence to justify doubts about an external world. The implicit assumption in classical skepticism, operative still, is that no experience epistemically grounds a belief if that experience is strictly compatible with an alternative belief. Whence the method of undermining belief by constructing alternatives.

B. An argument against empirical equivalence. We find the pervasiveness of this influence out of proportion to the conceptual credentials of the basic idea of empirical equivalence. By connecting three familiar and relatively uncontroversial theses, we can construct a simple argument to cast doubt on empirical equivalence in general, as a relation among scientific theories (and, by parity of reasoning, between any rival perspectives).

On the traditional view, theories are empirically equivalent just in case they have the same class of empirical, viz., observational, consequences.[2] A determination of empirical equivalence among theories therefore requires identifying their respective empirical consequence classes. As the empirical consequences of any statement are those of its logical consequences formulable in an observation language, these classes are (presumably proper) subsets of the logical consequence classes of theories. Central, therefore, to the standard notion of empirical equivalence are the notions of observational properties, the empirical consequences of a theory, and the logical consequences of a theory. We shall show that, when these concepts are properly understood, the doctrine of empirical equivalence loses all significance for epistemology.

Our three familiar theses are these:

Familiar thesis 1, the variability of the range of the observable (VRO):

> Any circumscription of the range of observable phenomena is relative to the state of scientific knowledge and the technological resources available for observation and detection.

---

[2] Empirical equivalence can also be formulated in semantic terms: empirically equivalent theories have the same class of empirical models. Although the point of the semantic approach is to achieve independence of theory from language, one still needs a criterion of empirical status or observability to formulate empirical equivalence. One also needs to circumscribe the class of models with which a theory is to be identified, and this will require some reference to the theory's axioms or basic assumptions. For the question whether or not some particular set of structures is a model for a theory is not answered definitionally in science. It is answered by attempting to *apply* the theory, by working out consequences. Collateral information is crucial to such applications For these reasons, we believe that questions about empirical equivalence must involve all the elements we shall bring to bear, whether the notion is formulated in ours or in semantic terms.

In particular, entities or processes originally introduced by theory frequently achieve observable or "empirical" status as experimental methods and instruments of detection improve. Such variability applies to any viable distinction between observational and theoretical language.[3]

Familiar thesis 2, the need for auxiliaries in prediction (NAP):

> Theoretical hypotheses typically require supplementation by auxiliary or collateral information for the derivation of observable consequences.

While direct derivability of statements bearing evidentially on theory is not in principle precluded, auxiliaries are generally required for the derivation of epistemically significant results.[4]

Familiar thesis 3, the instability of auxiliary assumptions (IAA).

> Auxiliary information providing premises for the derivation of observational consequences from theory is unstable in two respects: it is defeasible and it is augmentable.

Auxiliary assumptions once sufficiently secure to be used as premises frequently come subsequently to be rejected, and new auxiliaries permitting the derivation of additional observational consequences frequently become available.

Our argument against empirical equivalence now proceeds as follows. As VRO makes clear, the decision to locate a logical consequence of a theory outside its empirical consequence class (on the grounds of the former's nonobservational status) is subject to change. That class may increase, coming to incorporate an ever greater proportion of the theory's total consequence class.[5] This result already shows that findings of empirical equivalence are not reliably projectable, since we cannot reliably anticipate which of a theory's now unobservable consequences may become observable. But the problems with empirical equivalence run deeper than the inconstancy of the boundary of the observable. For even if it were possible to circumscribe the range of the observable relative to a state of science, we shall see that it would still be impossible so to circumscribe the range of auxiliary information available for use in deriving observational consequences.

---

[3] In labelling VRO "relatively uncontroversial," we acknowledge that van Fraassen's empiricism rejects it. Van Fraassen claims that what is observable is determined by facts about human beings as organisms, not by the transitory state of knowledge of those facts—not by science or technology. This view does not affect the use we shall make of VRO, however, because *judgments* of empirical equivalence must depend on *judgments* of what is observable; they cannot invoke transcendent facts.

But further, we reject the implicit assumption that conditions of observability are fixed by physiology Once it is decided what is to count as observing, physiology may determine what is observable. But physiology does not impose or delimit our concept of observation. We could possess the relevant physiological apparatus without possessing a concept of observation at all. The concept we do possess could perfectly well incorporate technological means of detection. In fact, the concept of observation has changed with science, and even to state that the (theory-independent) facts determine what is observable, van Fraassen must use a concept of observation that implicitly appeals to a state of science and technology.

[4] An appeal to NAP in criticism of empirical equivalence may appear ironic in view of the role of holistic theses in fostering belief in empirical equivalence and underdetermination. It is indeed ironic that the problems NAP, in concert with other theses, poses for empirical equivalence have escaped the notice of holists.

By NAP, a theory's empirical consequence class must be allowed to include statements deducible from the theory only with the help of auxiliaries. One can distinguish the broad from the narrow class of a theory's empirical consequences, where the narrow class contains only observational statements implied by the theory in isolation from other theories and hypotheses. But NAP shows that it is the broad class, containing as well statements deducible only if the theory is conjoined with such auxiliaries, that matters epistemologically. Regardless whether holists are right in contending that the narrow class is empty, it is a class of little epistemic moment. It is by the complement of the narrow with respect to the broad that theories are primarily tested, and a characterization of empirical equivalence limited to the narrow would have no such epistemological consequences as we are concerned to contest.

It follows by IAA that, apart from shifts in observational status, a theory's empirical consequence class may increase through augmentations to the theory's total consequence class. As new auxiliary information becomes available, new empirical consequences derived with its help are added. Of course, conditionals connecting the auxiliary statements newly used to the empirical statements newly derived were already present among the theory's logical consequences. But the detached empirical statements are not present until the auxiliaries on which their deducibility depends become available. So long as we include within a theory's empirical consequence class statements derivable from the theory only via auxiliaries, so long as we construe that class *broadly*—and we have argued that it must be so construed to reflect the realities of theory testing—the theory's logical consequence class will be augmentable in virtue of containing the empirical consequence class as a subset. The empirical consequence class can also diminish, again by IAA, as the rejection of needed auxiliaries discontinues the derivability of some of its members. Therefore, any determination of the empirical consequence class of a theory must be relativized to a particular state of science. We infer that empirical equivalence itself must be so relativized, and, accordingly, that any finding of empirical equivalence is both contextual and defeasible.

---

[5] Conceivably, it could also decrease through a shift in the status of a consequence from observational to nonobservational, although that is not the usual pattern in the sciences.

This contextuality shows that determinations of empirical equivalence are not a purely formal, a priori matter, but must defer, in part, to scientific practice. It undercuts any formalistic program to delimit the scope of scientific knowledge by reason of empirical equivalence, thereby defeating the epistemically otiose morals that empirical equivalence has been made to serve. The limitations on theoretical understanding that a defeasible empirical equivalence imposes need not be grievous. Nevertheless, we think there is still less to the notion of empirical equivalence than survives these concessions.

It has been widely supposed that one can, utilizing the resources of logic and semantics alone, "read off" the observable consequences of a theory. The mobility of the boundary of the observable has been regarded as an inconvenience, not as a fundamental challenge to the idea that, at least in principle, the consequences or content of a theory are unambiguously identifiable. On this view, enhancements to our observational repertoire do nothing to alter a theory's semantics; rather, they merely shift the line, within the class of a theory's logical consequences, between observational and nonobservable consequences. But NAP shows that there is an epistemic question here quite distinct from the logico-semantic one. Specifically, the availability of auxiliaries—auxiliaries crucial for determining what a theory's empirical consequences are—is neither a matter of logic nor semantics; it is inescapably epistemic. The determination that a given empirical statement, *e*, is an empirical consequence of a particular theory, T, depends on whether there are epistemically well-grounded collateral hypotheses that establish a suitable inferential link between T and *e*. The availability of such hypotheses is clearly a matter of evidential warrant. Once statements whose derivability requires auxiliaries are allowed to count as consequences, as they must by NAP, no statement can be disallowed as a (broad) consequence of a theory unless some statements are disallowed as auxiliaries. So, before deciding whether a derivation of an observation statement from a theory plus auxiliaries qualifies the statement for inclusion in the theory's empirical consequence class, we must assess the epistemic standing of the auxiliaries.

This makes clear that the epistemic bearing of evidence on theory is not a purely logical relation, but is subject to reinterpretation as science grows and may be indeterminate at a particular point in the process of growth. How well supported an auxiliary is by evidence available *now* may depend on findings made *later*—a problem exacerbated by the fact that standards of evidential support themselves are transformable by the fortunes of empirical beliefs.[6] This fact darkens the prospects even for time-indexed delineations of empirical consequence classes.

C. Response to anticipated objections. The response we anticipate to our argument is a challenge to its assumption that empirical consequence classes must be identified for their equivalence to be established. Can there not be a general argument to show that classes must be the same independently of determining their membership? An obvious suggestion is that logically or conceptually equivalent theories must have the same consequence class, whatever that class is. As we do not question the empirical equivalence of logically equivalent theories, we ignore this suggestion and assume henceforth that theories whose empirical equivalence is at issue are logically and conceptually distinct.

One approach to constructing a general argument is to invoke the Lowenheim-Skolem Theorem. This theorem asserts that any first-order, formal theory that has a model at all has a denumerable model. A standard proof uses terms involving individual constants indexed by the natural numbers as the domain of a model. But if the domain need only be a set of terms, it could just as well be any denumerable set whose members are proposed as the referents of those terms. So, in principle, such a theory has an infinite number of models.

The qualification 'in principle' must be emphasized here, because there is no guarantee that the denumerable models of a consistent, formal theory are effectively constructable. The proof of the theorem relies on the completability of a consistent first-order theory, and the complete theory need not be axiomatic even if the theory it completes is axiomatic. Thus, any appeal to the Lowenheim-Skolem Theorem on behalf of empirical equivalence shows at most that equivalent theories exist in principle; it does not show that they are entertainable as alternatives.

But there is a more fundamental objection to the relevance of the theorem. Having multiple models of a formal theory does not mean having multiple theories of common empirical content. A physical theory, by virtue of being a physical theory, includes a semantic interpretation of its formal structure; it is not simply a formal structure variously interpretable. A physical theory is inherently at least purportedly referential. Its referents may turn out not to exist, if nature fails to cooperate. But what its referents are if it has them is fixed by the theory itself; it is not a matter of optional interpretation. The reference-fixing devices of physical theory are rich and various. It does not require a philosophical theory of reference fixing to make the point that the semantic resources of physical theories transcend syntactic requirements on first-order formal theories. If, given the Lowenheim-Skolem Theorem, formal statements in first-order logic are referentially indeterminate, then a physical theory is not simply a set of formal statements in first-order logic.

---

[6] See our contributions to the symposium, "Normative Versions of Naturalized Epistemology," *Philosophy of Science*, LVII, 1 (1990): 20–34; 44–60

Another approach is to construct an algorithm for generating empirical equivalents to a given physical theory, such as the Lowenheim-Skolem Theorem fails to do for formal theories. For example, there exist instrumentalist algorithms for excising the theoretical terms of a theory without empirical loss. Whether such algorithms are in fact successful is rendered highly dubious by the premises of our argument. It is by no means clear that a theory's instrumentalized version can match its capacity for empirical commitment, once the role of auxiliaries in fixing such commitment and the variability of the range of the observable are acknowledged. At most a theory's instrumentalized version can be held empirically equivalent to it relative to a circumscription of the observable and a presumed or intended domain of application. But while theories fix their own intended interpretations, they do not fix their own domains of application, nor the resources for detection of entities they posit. Algorithmically excised references may pick out entities that become detectable. New applications may arise with changes in collateral knowledge. Indeed, it is a measure of a theory's success when posited entities acquire a technological role, and applications for which the theory was not designed become possible.[7]

Be that as it may, what application of an instrumentalist algorithm to a theory produces is manifestly not an alternative *theory*. That is, the algorithm does not produce a rival representation of the world from which the same empirical phenomena may be explained and predicted. On the contrary, a theory's instrumentalized version posits nothing not posited by the theory, and its explanations, if any, of empirical phenomena deducible from it are wholly parasitic on the theory's own explanations. A theory's instrumentalized version cannot be a rival to it, because it is a logical consequence of the theory and is bound to be endorsed by anyone endorsing the theory. The challenge the instrumentalist poses is to justify endorsing more than the instrumentalized version, not to justify endorsing something instead of it. The point of the instrumentalist move is not to challenge the epistemic status of theory by demonstrating underdetermination among alternatives, but to argue that empirical evidence underdetermines theories individually by failing to discriminate between them and their instrumentalized versions. We know of no algorithm for generating genuine theoretical competitors to a given theory.[8]

---

[7] Consider the stature gained by Newtonian mechanics through its unforeseen applicability to the motions of fluids, and unforeseeable applicability to electric current.

The only other approach we know of to establishing empirical equivalence without identifying empirical content is to argue from cases. We propose an example, inspired by van Fraassen's in *The Scientific Image*, as representative. Let TN be Newtonian theory. Let R be the hypothesis that the center of mass of the expanding universe is at rest in absolute space. Let V be the hypothesis that the center of mass of the universe has constant absolute velocity v. Consider the claim that TN + R is empirically equivalent to TN + V.

This claim is based on the common TN component of the theories. It is Newtonian theory itself that assures us that unaccelerated absolute motion has no empirical consequences of a kind encompassed by the theory; that is, no consequences within mechanics. We can therefore bring two lines of criticism against the claim of empirical equivalence: either there is some other kind of consequence not envisioned within mechanics, or the underlying Newtonian assurance is wrong. The question is whether conceivable developments in scientific knowledge enable us to distinguish the theories empirically on one of these bases.

Van Fraassen's defense of his claim of empirical equivalence does not take account of these possibilities. He considers only extensions of the theories to further, late-breaking mechanical phenomena, arguing that whatever these may be, further equivalent extensions are constructable (*op. cit.*, p. 47ff). Can a defense be provided to cover the possibilities that van Fraassen neglects?

---

[8] We discount here the trivial algorithm that, applied to any theory T committed to theoretical entities of type r, generates the "rival" T* that asserts the world to be observationally exactly as if T were true but denies the existence of rs. Its logical incompatibility with T is insufficient to qualify T* as a genuine rival, as T* offers no competing explanations and is totally parasitic on T for whatever virtues it does offer For extended criticism of such devices, see Leplin, "Surrealism," *Mind*, XCVI (October 1987): 519–24.

Imagine that TN + V has nonmechanical consequences absent from TN + R. A new particle, "the bason," is independently hypothesized to arise with absolute motion. The positive absolute velocity of the universe represents energy available for bason creation, and basons will appear under certain conditions in principle realizable in the laboratory if TN + V is correct, but not if TN + R is correct. v can be measured by counting basons, so that variants of TN + V for different v are empirically distinguishable.

We can construct an extension of TN + V which agrees with TN +R in not predicting basons. Let TN + W be TN + V plus the hypothesis that there is a velocity w such that basons appear if and only if and to the extent that $v > w$. Then the absence of basons establishes only that v does not exceed w; it does not require R. The presence of basons still refutes TN + R, but TN + R can be supplemented to allow basons; perhaps they arise spontaneously. TN + R then lacks an explanation of bason production, such as TN + V provides. Something in the way of explanatory parity is achievable by adding to TN + R the hypothesis that what absolute motion produces is antibasons, which immediately annihilate basons. So the presence of basons is explained by the lack of absolute velocity. Still, TN + R does not explain the frequency of bason detection, as TN + V does. The observed frequency must simply be posited, as a constant determined by experiment, and this procedure is an admitted disadvantage relative to TN + V. But this comparison does not affect empirical equivalence.

The appeal to nonmechanical, differentiating phenomena can be defeated, because, if empirical equivalence holds within mechanics, it continues to hold for any extensions of mechanics in which the presence or absence of additional, nonmechanical phenomena is made to depend on the value of a mechanical property. This seems to be a general result. If theories T_1 and T_2 are equivalent with respect to properties $p_1, \ldots, {}_n$, they have equivalent extensions for any enlarged class of properties $p_1, \ldots, {}_n, q_1, \ldots, {}_m$; where properties $q_1, \ldots, {}_m$ are functions of $p_1, \ldots, {}_n$. On the other hand, if $q_1, \ldots, {}_m$ are not functions of $p_1, \ldots, {}_n$, they cannot be used to discriminate between T_1 and T_2.

It remains to consider the common core TN of the allegedly equivalent theories of the example. Here our argument finds its proper vindication. The point to make is that this core, and with it the original rationale for regarding TN + R and TN + V as equivalent in the first place (that is, unextended to nonmechanical properties) is defeasible. We may have good or sufficient reason to regard theories as empirically equivalent, but there is no guarantee. That concession is all our argument requires. We do not deny the possibility that the world is such that equally viable, incompatible theories of it are possible. We do not deny the possibility of the world's being unamenable to epistemic investigation and adjudication, beyond a certain level. But whether or not the world is like that is itself an empirical question open to investigation. The answer cannot be preordained by a transcendent, epistemic skepticism.

It is noteworthy that contrived examples alleging empirical equivalence always invoke the relativity of motion; it is the impossibility of distinguishing apparent from absolute motion to which they owe their plausibility. This is also the problem in the pre-eminent historical examples, the competition between Ptolemy and Copernicus, which created the idea of empirical equivalence in the first place, and that between Einstein and H. A. Lorentz.[9] Either the relativity of motion is a physical discovery, founded on such evidence as the empirical success of Newtonian or relativity theory, or it is guaranteed conceptually. Both views have been advanced historically. Our reply to alleged examples of empirical equivalence takes the former view. We maintain not that there are no cases of empirical equivalence, but that the claim that there are is defeasible.[10] If the latter view is to be assumed, then we return to the caveat that we are not concerned to contest the empirical equivalence of conceptually equivalent theories.

II. UNDERDETERMINATION

We have argued that the thesis that every empirically successful theory has empirically equivalent counterparts is precarious, at best. But for now let us suspend our incredulity about empirical equivalence and suppose that the thesis is sound. We wish to explore in this section what, if anything, then follows from the existence of empirically equivalent theories for general epistemology.

A number of deep epistemic implications, roughly collectable under the notion of "underdetermination," have been alleged for empirical equivalence. For instance, it is typical of recent empiricism to hold that evidence bearing on a theory, however broad and supportive, is impotent to single out that theory for acceptance, because of the availability or possibility of equally supported rivals. Instrumentalists argue that the existence of theoretically noncommittal equivalents for theories positing unobservable entities establishes the epistemic impropriety of deep-structure theorizing, and with it the failure of scientific realism. Some pragmatists infer that only nonepistemic dimensions of appraisal are applicable to theories, and that, accordingly, theory endorsement is not exclusive nor, necessarily, even preferential. One may pick and choose freely among theories whatever works for the problems at hand, so that the distinction between theories and models is lost. In a phrase, the thesis of underdetermination, denying the possibility of adequate evidential warrant for any theory, has become the epistemic corollary to the presumptively semantic thesis of empirical equivalence.

---

[9] Contemporary examples are also limited to space-time theories, raising the possibility of underdetermination for certain topological features of space-time. For example, it might be possible to obtain the same consequences from a dense but discreet space-time as are obtained by adding dimensions to continuous space-time. A more general treatment of relative motion would subsume it under topological considerations.

[10] And we deny the omnibus a priori claim that every theory has empirically equivalent rivals.

Against these positions, we shall argue that underdetermination does not in general obtain, not even under conditions of empirical equivalence. As we have seen, empirical equivalence is chiefly seen as a thesis about the *semantics* of theories; underdetermination, by contrast, is a thesis about the *epistemology* of theories. It has been supposed that, if theories possess the same empirical consequences, then they will inevitably be equally well (or ill) supported by those instances. We shall contest this supposition and, with it, the reduction of evidential relations to semantic relations, on which it rests. We dispute the ability of semantic considerations to resolve epistemic issues. But even allowing the epistemic dimension we have discerned in empirical equivalence, we shall find that the relative degree of evidential support for theories is not fixed by their empirical equivalence.

Specifically, we shall show, first, that significant evidential support may be provided a theory by results that are not empirical consequences of the theory;[11] secondly, we shall show that (even) true empirical consequences need lend no evidential support to a theory. By this dual strategy, we propose to sever the presumed link between supporting instances and empirical consequences.[12] We shall show that being an empirical consequence of a hypothesis is neither necessary nor sufficient for being evidentially relevant to a hypothesis. These conclusions will establish that theories identical as to empirical consequences may be differentially supported, such that one is epistemically preferable to the other.

---

[11] Here we allow, as always, a role for auxiliaries in generating consequences.

[12] In "Realism, Underdetermination, and a Causal Theory of Evidence," *Nous*, VII (March 1973) 1–12. Richard Boyd sought to drive a wedge between empirical consequences and supporting instances But Boyd's case rests on the (in our view dubious) principle that "new theories should, prima facie, resemble current theories with respect to their accounts of causal relations among theoretical entities" (p. 8). Our argument will require no such restriction.

A. *Evidential results that are not consequences.* We begin by noting that instances of a generalization may evidentially support one another, although they are not consequences of one another. Previous sightings of black crows support the hypothesis that the next crow to be sighted will be black, although that hypothesis implies nothing about other crows. Supposing this evidential connection to be uncontroversial, we ask why, then, in the case of universal statements it should be supposed that evidential support is limited to logical consequences.

Is it that the evidential connection admitted to hold among singular statements is at best indirect, that it connects those statements only via a general statement that they instantiate? The thesis would then be that *direct* evidential support for a statement is limited to its logical consequences, and singular statements instantiating the same generalization support one another only in virtue of directly supporting that generalization. In short, where there appears to be evidential support for a statement, *s*, outside the range of *s*'s logical consequences, such support is parasitic on support of a general statement, *m*, which entails *s*, from *m*'s logical consequences.

We believe this to be an unperspicacious way of accounting for what goes on in singular inference. Often the evidential link between singular statements is stronger than the support available for a general intermediary, whose identification can, in any case, prove elusive. But even if this account worked, it should be noted straightaway that allowing a statement to accrue indirect empirical support in this fashion already undermines the claim that statements are confirmable only by their empirical consequences. This result alone suffices to establish that the class of empirical consequences of a statement and the class of its prospective confirming instances are distinct.

We began this discussion with the hackneyed case of black crows in order to show that the possibility of inferences of even the most mundane sort (from particular-to-particular) depend upon denying the thesis that evidential support accrues to a statement only via its positive instances. This claim becomes even clearer when one considers the manner in which real scientific theories garner empirical support. Consider, for instance, the theory of continental drift. It holds that every region of the earth's surface has occupied both latitudes and longitudes significantly different from those it now occupies. It is thereby committed to two general hypotheses:

H_1: There has been significant climatic variation throughout the earth, the current climate of all regions differing from their climates in former times.

H_2: The current alignment with the earth's magnetic pole of the magnetism of iron-bearing rock in any given region of the earth differs significantly from the alignment of the region's magnetic rocks from earlier periods.

During the 1950s and 1960s, impressive evidence from studies of remnant magnetism accumulated for H_2. Clearly, those data support H_1 as well, despite the fact that they are not consequences of H_1. Rather, by supporting H_2 they confirm the general drift theory, and thereby its consequence H_1.

Similar examples are readily adduced. Brownian motion supported the atomic theory—indeed, it was generally taken to demonstrate the existence of atoms—by being shown to support statistical mechanics. But, of course, Brownian motion is no consequence of atomic theory. The increase of mass with velocity, when achieved technologically in the 1920s, supported kinematic laws of relativity which to that point continued to be regarded with great suspicion. J. J. Thompson's cathode ray experiments in the 1890s were important evidence for a host of theoretical hypotheses about electricity that depended on Lorentz's electroatomism. Phenomena of heat radiation were used by J. C. Maxwell in the 1870s to support the kinetic molecular theory, which did not address the transmission of heat energy across the space intervening between bodies. The emergence in the 1920s of evidence showing heritable variation supported Darwin's hypothesis about the antiquity of the earth, although that hypothesis entailed nothing about biological variation. Contemporary observational astronomy is replete with indirect methods of calculating stellar distances, whereby general hypotheses in cosmology acquire support from facts they do not imply about internal compositions of stars.

A number of points are to be noted about such examples. First, by dating them we emphasize that they are not dismissable by invoking auxiliaries via which the evidence is derivable. One could not in the 1890s represent Thompson's results as consequences of electrical laws by making electroatomism an auxiliary. Despite Ludwig Boltzmann's pioneering work, statistical mechanics was too speculative in 1905 to qualify as an available auxiliary. Even taking an ahistorical view, it would be casuistical to represent evidence as a consequence of a hypothesis from which it is derivable via auxiliaries, if it is the auxiliaries rather than the hypothesis that really fuel the derivation. If a formal criterion is wanted, we may stipulate that a hypothesis be ineliminable from the derivation of what are to qualify as its consequences.[13]

Second, the more general theory via which the evidence supports a hypothesis of which it is not a consequence need not be very precise or specific. For example, the statistical mechanics that Brownian motion supported was more a program for interpreting phenomenological thermodynamics probabilistically than a developed theory. There can be good reason to believe that conceptually dissimilar hypotheses are related such that evidence for one supports the other, without possessing a well worked out or independently viable theory that connects them. Perhaps a theory that connected them has been discredited without the connection it effected being discredited. In this respect, nonconsequential evidence for general statements approximates the case of singular statements for which the inferential link proved elusive.

Third, we need not fear running afoul of familiar paradoxes of confirmation in taking evidence to confirm a hypothesis in virtue of supporting a more general statement that implies the hypothesis. The intuition that what increases our confidence in a statement thereby increases our confidence in what that statement entails is fundamentally sound. The difficulties that Carl Hempel, for example, extracted from his "special consequence condition" depend on a certain logical form for general laws and a simplistic criterion of confirmation—Nicod's criterion—which requires, in opposition to the position we have undertaken to defend (see section II.B), that all positive consequences be confirming. Much sophisticated reasoning in the natural sciences would be vitiated by restricting evidence relevant in assessing a theory to the entailments (via auxiliaries) of the theory. And any singular prediction would be so vitiated as well.

Finally, we need to acknowledge and take into account a subtlety of confirmation that might appear to challenge the force of nonconsequential evidence for our argument. There is an obvious way in which a statement not entailed by a theory can be evidence for the theory. The statement might imply another empirical statement that *is* entailed. Suppose, for example, that the theory entails a—perhaps indefinitely extendable—disjunction of which the statement is a disjunct. By implying a statement that is a consequence, the evidence, though not itself a consequence, fails to discriminate between the theory and any empirically equivalent theory. So showing that there can be evidence for a theory that is not a consequence of the theory does not suffice to show that empirically equivalent theories can be differentially supported.

---

[13] Logico-semantic trickery will not render it ineliminable so long as epistemic conditions are imposed on auxiliaries.

Our examples, however, do not pose this problem. An example that does is to cite ten successive heads in tosses of a coin as evidence that the coin is biased. The hypothesis of bias does not entail the results of any given number of trials. But it does entail that something such as an improbably extended succession of heads will result over many trials. A hypothesis empirically equivalent to that of bias, were there one, would entail the same thing, and thereby be supported by any evidence thus exemplifying bias. On the other hand, the hypothesis of bias readily admits of evidential support from sources outside its consequence class that would not support purportedly equivalent hypotheses. An example is the information that the coin hypothesized to be biased was poured in a die cut by a chronically inebriated diemaker.

What, then, *is* the connection that we claim our examples to establish between nonconsequential evidence and differential support of empirically equivalent theories? We propose the following exemplar. Theoretical hypotheses H_1 and H_2 are empirically equivalent but conceptually distinct. H_1, but not H_2, is derivable from a more general theory T, which also entails another hypothesis H. An empirical consequence e of H is obtained. e supports H and thereby T. Thus, e provides indirect evidential warrant for H_1, of which it is not a consequence, without affecting the credentials of H_2. Thus, one of two empirically equivalent hypotheses or theories can be evidentially supported to the exclusion of the other by being incorporated into an independently supported, more general theory that does not support the other, although it does predict all the empirical consequences of the other. The view that assimilates evidence to consequences cannot, on pain of incoherence, accept the intuitive, uncontroversial principle that evidential support flows "downward" across the entailment relation.

We stress, however, that this is *only* an exemplar. There are modes of nonconsequential empirical support in science that do not invoke intermediate theories or generalizations. This has already been seen in the case of singular inference, and is present as well in the example from astronomy. Another type of example is the use of analogical reasoning. Analogical reasoning is often motivational, pertaining more to the heuristics of theory development than to confirmation. But sophisticated analogies can be evidentially probative.

Maxwell analogized a closed system of elastic particles to a contained gas, whereby the mathematical theory of collisions together with the observed properties of gases supports a molecular structure for gas. The power of the analogy to yield important, known properties of gases makes it reasonable to infer, Maxwell[14] thinks, that "minute parts" of gases are in rapid motion. Einstein supported his hypothesis of a quantum structure for radiation by analogizing the entropy of monochromatic radiation to the entropy of an ideal gas. He showed that a decrease in entropy corresponding to a reduction in volume for radiation of fixed energy has the same functional form as the decrease in entropy associated with a reduction in volume for an ideal gas. Evidence that warrants treating the gas statistically thereby supports a quantum structure for the gas. Yet such evidence sustains no logical relation to Einstein's hypothesis, nor does it support any theory that entails Einstein's hypothesis.

The many examples we have adduced may seem too obvious and frequent to dwell on. We welcome that reaction, but admonish many influential epistemological positions of the last half century for ignoring them. Karl Popper, for instance, presumes throughout his treatment of scientific methodology that evidence potentially relevant to the assessment of a hypothesis must come from its empirical consequence class. And many epistemologists follow Quine in supposing that the empirical equivalence of rival hypotheses renders inevitable their epistemic parity, reducing a choice among them to purely pragmatic considerations. In defiance of much celebrated epistemology, we claim that our examples establish that theories with exactly the same empirical consequences may admit of differing degrees of evidential support.

B. Empirical consequences that are not evidential. Establishing that evidential results need not be consequences is already enough to block the inference from empirical equivalence to underdetermination. But it is instructive to make the converse point as well. Suppose a televangelist recommends regular reading of scripture to induce puberty in young males. As evidence for his hypothesis (H) that such readings are efficacious, he cites a longitudinal study of 1000 males in Lynchburg, Virginia, who from the age of seven years were forced to read scripture for nine years. Medical examinations after nine years definitively established that all the subjects were pubescent by age sixteen. The putatively evidential statements supplied by the examinations are positive instances of H. But no one other than a resident of Lynchburg, or the like-minded, is likely to grant that the results support H.

---

[14] "Illustrations of the Dynamical Theory of Gases," in *The Scientific Papers of James Clerk Maxwell*, W. D. Niven, ed. (New York. Cambridge, 1890).

This example has a self-serving aspect. That the televangelist has a pro-attitude toward H on grounds independent of the purported evidence he cites is already enough to make one wary; one need not recognize the flaws in the experimental design of the longitudinal study. In a case without this feature, a person hypothesizes that coffee is effective as a remedy for the common cold, having been convinced by finding that colds dissipate after several days of drinking coffee. The point here is that the very idea of experimental controls arises only because we recognize independently that empirical consequences need not be evidential; we recognize independently the need for additional conditions on evidence.

No philosopher of science is willing to grant evidential status to a result *e* with respect to a hypothesis H just because *e* is a consequence of H. That is the point of two centuries of debate over such issues as the independence of *e*, the purpose for which H was introduced, the additional uses to which H may be put, the relation of H to other theories, and so forth.

Results that test a theory and results that are obtainable as empirical consequences of the theory constitute partially nonoverlapping sets. Being an empirical consequence of a theory is neither necessary nor sufficient to qualify a statement as providing evidential support for the theory. Because of this, it is illegitimate to infer from the empirical equivalence of theories that they will fare equally in the face of any possible or conceivable evidence. The thesis of underdetermination, at least in so far as it is founded on presumptions about the possibility of empirical equivalence for theories—or "systems of the world"—stands refuted.

III. FORMAL CONSTRAINTS ON EPISTEMOLOGY

If the identification of empirical consequences with evidential support is so implausible, how has it managed to gain such a foothold? We suggest that a more persuasive, less readily dispelled confusion is ultimately responsible. That confusion, as we have intimated, is to misunderstand the relationship between semantics and epistemology, bringing the largely technical and formal machinery of semantics improperly to bear on epistemic issues.

Specifically, we wish to reveal and challenge the widespread—if usually implicit—conviction that epistemic relations are reducible to semantic relations. It is commonly supposed either that truth and meaning conditions just *are* justification conditions, or, at least, that they can be made to double as justification conditions.[15] Either way, epistemology is made the poor relation of a family of interconnections among semantic, syntactic, and epistemic concepts, and is left to make do with tools handed down from semantics. It seems to us that distinctively epistemic issues are left unresolved by such a presumed reduction, and that epistemic theses depending on it—such as the underdetermination thesis—are wrongheaded. We will first explain and illustrate the confusion we have diagnosed, then trace the mistaken assimilation of support to empirical consequences to it.

---

[15] We do not mean to suggest that all epistemologists are guilty of conflating epistemic and semantic relations. But there are some influential theorists of knowledge, e.g., Quine, who do precisely this.

The problem originates in foundationalist epistemology—especially in Descartes's image of a mathematically rigorous, deductive structure for knowledge—and thus is not confined to empiricism. If the evidential relation is deductive, the evidence on which a knowledge claim is based must bear semantic relations to the claim sufficient to permit the deduction.

Perhaps the best modern illustration is the attempt by the logical empiricists to demarcate science by semantic means. Both "verifiability" and "falsifiability," the prime concepts in terms of which scientific status has been delimited, are tools of semantics. Demarcation criteria proposed by the Vienna Circle or its positivist disciples, and by the Popperians, are alike in depending on semantic analysis and syntactic form of statements. What is required for classification as verifiable or falsifiable is basically that a statement satisfy constraints as to logical form and be couchable in observation language. It was assumed on all sides that such conditions suffice to identify the class of statements that are properly the objects of scientific inquiry.

The problem that demarcation criteria were intended and offered to solve, however, was the epistemic one of judging the reasonableness of belief. What was called for was a distinction, impossible within semantics alone, between statements that are well-founded and those that are not. Whether one's target was the metaphysics, religion, or ethics that exercised the positivist, or the psychoanalysis, Marxism, or astrology that exercised the Popperian, it was the irrationality of credence that the target was to be convicted of, not impropriety of logical form or visual inaccessibility of subject matter.

One might think to defend the adequacy of semantic tools for the intended distinction by arguing that the relevant notion of "science" to be demarcated is not that of what passes muster by scientific standards, but merely that of what is up for grabs in scientific inquiry. After all, statements falsified by scientific inquiry are yet to be classed as scientific. But not only is a distinction between what *qualifies* as scientific and what does not basically epistemic; so too is a distinction between what is *worthy* of investigation or *entertainable* by scientific means and what is not. It is basically what we have already found it reasonable to believe that decides these things.

The demarcation problem of the logical empiricists arose as a variant on the logical positivists' program for distinguishing cognitive significance from emotive uses of language misleadingly given propositional form. Already at this level one may discern the assimilation of evidential to semantic relations. For the evaluative force of, e.g., ethical pronouncements that led positivists to disqualify them as genuine propositions is also present in epistemic pronouncements, and, derivatively, in science. Epistemology, it is now commonly recognized, is value-laden.[16] But science was the logical positivists' paradigm of cognitive significance; its propositional status, the ideal to which ethics, religion, and metaphysics futilely aspired. If epistemology, and science in particular, was to be salvaged, then epistemic evaluation would have to rest on semantic relations as the only factual alternative to value-free empirical relations.

Accounts of scientific explanation and confirmation proposed in the 1940s and 1950s exhibit the same priority of the semantic over the epistemic. Characteristically, these accounts dealt incidently if at all with epistemic and pragmatic dimensions of confirmation and explanation, in favor of their syntax, logical structure, and semantics. Significantly, truth enters as a precondition of explanatory status, rather than as an attribute that it is rational to adduce (partly) in measure that explanations are achieved.

What made such approaches seem plausible, we suggest, was a linguistic view of conceptual analysis. To analyze a concept, one examined its use in language. To understand instances of its use in language, one identified the truth conditions for sentences. In particular, to analyze knowledge one identified the truth conditions for attributions of knowledge; for sentences of the form ". . . knows . . . ." It turns out, however, that among the truth conditions for such a sentence is the truth of another sentence, that disquoted in the second position of the schema. It therefore looks like the semantic concept of truth is logically prior to the epistemic concept of knowledge.

Of course, this does not make the semantic concept sufficient for the epistemic one, but further developments tended to elevate semantics and syntax over the notions of evidential warrant and rationality of belief used in other truth conditions for knowledge attributions. The incompleteness of the list of truth conditions was manifested in a curious asymmetry between the truth and evidence conditions. If the truth condition is not met, no bolstering of the evidence is sufficient for knowing. But inconclusive evidence that leaves open the possibility of error can be sufficient for knowing, if only, as a matter of fact (or happenstance), the world cooperates. In many celebrated paradigms of knowing, the evidence needed does not seem all that strong.[17] Thus, attention focused more on the truth condition than the evidence condition—more on semantic than epistemic issues. Ironically, the recent emergence of reliability theory, which re-emphasizes the justificatory component of knowledge in the tradition of Gettier's challenge, underscores the paucity and defeasibility of the evidence on which ordinary knowledge relies. Add to this asymmetry the success of Tarski's theory of truth in contrast to the sorry state of theories of evidential warrant, and one has the makings of a semantic and syntactic orientation for epistemology.

---

[16] See Lycan, "Epistemic Value," *Synthese*, LXIV (1985): 137–64.

Given this orientation, it was natural to approach the problem of warranting a hypothesis—the problem of testing—by attending to statements that bear syntactic and semantic relations to the hypothesis—to its instantiations. At least this was natural for empirical generalizations, whose instantiations are empirical statements. This approach then created so many internal problems and tasks—Hempel's paradoxes of confirmation across logical relations; Goodman's problem of projectability—that the possibility of warrant provided by statements syntactically and semantically independent of the hypothesis was lost sight of. Instantiations of theoretical hypotheses are not empirical, but an assimilation of support to consequences was somehow extrapolated for them, by supposing them in principle recastable in observational terms or, perhaps, by supposing their testability reducible to the testability of empirical generalizations. Such was the hold of the resulting picture, that the assimilation of support to consequences exceeded the confines of logical empiricism to capture the format of textbook characterizations of scientific method itself. Although written by a philosopher, Hempel,[18] the following passage will strike every reader as stereotypical of standard accounts of empirical inquiry:

> First, from the hypothesis under test, suitable other statements are inferred which describe certain directly observable phenomena that should be found to occur under specifiable circumstances if the hy-

---

[17] This is a striking feature of Gilbert Harmon's examples in *Thought* (Princeton. University Press, 1973).

[18] *Aspects of Scientific Explanation* (New York: Free Press, 1965).

> pothesis is true; then those inferred statements are tested directly, i.e., by checking whether the specified phenomena do in fact occur; finally, the proposed hypothesis is accepted or rejected in the light of the outcomes of those tests (*ibid.*, p. 83).

This is a pure and simple statement of a view that deserves to be called "consequentialism," viz., the thesis that hypotheses are to be tested exclusively by an exploration of the truth status of those empirically decidable statements which they entail. Consequentialism is closely connected historically with the ancient idea of "saving the phenomena," that is, of reconciling theory with aberrant phenomena by introducing auxiliaries permitting those phenomena to be derived. Although not the invention of empiricists, it held a special appeal for them in appearing to solve problems of meaning, truth, and justification in one go. Meaning conditions, truth conditions, and justification conditions became substantially the same thing. The pre-Tarskian slogan, "To know what *x* means is to know how to test *x*," became, after Tarski, "The truth conditions for *x* specify simultaneously the meaning of *x* and the test conditions for *x*." [19] And this cast of mind, we suggest, is behind the supposition that statements with the same empirical consequences must be on an equivalent epistemological footing.

It is remarkable, upon reflection, how much freight the entailment relation has been made to carry in recent philosophy. The early logical positivists identified the meaning of a statement with what that statement entailed in the observation language. The logical empiricists and Popper sought to demarcate scientific from nonscientific statements solely by whether or not they entailed observational statements (Rudolf Carnap) or negations of observation statements (Popper). The Tarskian account of truth held that a theory was true just in case every statement constituting its logical content was true, and in practice a statement's logical content was to be fixed by its entailments.

In twentieth-century epistemology, we find the notorious Nicod criterion, discussed at length by Hempel and other confirmation theorists of the 1950s and 1960s, presuming to define the evidential relation exclusively in terms of positive and negative instances of the hypothesis, where positive instances instantiate the hypothesis and negative instances instantiate its negation. Popper's influential account of theory testing held that genuine tests of a theory had to be drawn from statements in its empirical content class, a class defined as the set of empirical statements whose negations were entailed by the theory. Currently, van Fraassen's constructive empiricism defines its key evaluative concept—empirical adequacy—in terms of the observational structures that model a theory:

---

[19] Compare Quine: ". . . epistemology now becomes semantics. For epistemology remains centered as always on evidence, and meaning remains centered as always on verification; and evidence is verification"—"Epistemology Naturalized," p. 89.

> . . . a theory is empirically adequate exactly if what it says about the observable things and events in this world, is true—exactly, if it 'saves the phenomena' (*op. cit.*, p. 12).

And that a structure models a theory is determined not, as we have remarked, definitionally, but by reference to the theory's entailments. Clark Glymour, an avowed scientific realist, develops a "boot-strapping" story about theory testing, which insists that the statements capable of testing a hypothesis imply its empirical consequences. And Quine repeatedly avers that the only central rule of scientific method is hypothetico-deduction.

This ubiquitous assimilation of a theory's test cases to its logical consequences in an observation language, as we have argued above, wrongly ignores some of the more salient ways of testing theories. Worse, it generously greases the slide from empirical equivalence to underdetermination and epistemic parity. Ironically, the limitation of a statement's justification conditions to its truth conditions represents a striking break with the traditional empiricist project. Prior to the emergence of neopositivism in the 1920s, the general idea about theory testing and evaluation was that there was a range of "phenomena" for which any theory in a particular field was epistemically accountable. (In planetary astronomy, for example, these phenomena would be observations of positions of the planets, sun, and moon.) A theory's success or failure was measured against these phenomena, and decided by the theory's ability to give an account of them. A theory was, of course, responsible for its entailments, but it was held equally accountable for all the relevant, established phenomena, and could not evade this responsibility by failing to address them. For a Newton, a Ptolemy, or a Mach, "saving the phenomena" meant being able to explain all the salient facts in the relevant domain.[20]

With the rise of neopositivism, the epistemic responsibilities of theories were radically reinterpreted. Theories became liable only for what they entailed. Failure to address relevant phenomena, or at least to be indirectly applicable to them, now emerges as a cheap way of protecting such success as a theory does achieve, rather than as a liability. Where empirical adequacy formerly meant the ability to explain and predict all the salient phenomena, it now requires only possession of none but true empirical consequences. Recall the passage lately quoted from van Fraassen. The radical character of the shift we are describing becomes immediately clear there when one notes his identification of "empirical adequacy," saying only true things about observable features of the world, and "saving the phenomena." Prior to our time, no one would have supposed, as does van Fraassen, that saving the phenomena amounts only to possessing an observable model. No one would have supposed, as does van Fraassen, that a theory is to be judged only against the correctness of its own observational commitments (be those commitments expressed in model-theoretic or propositional form), irrespective of the comprehensiveness of the class of such commitments, irrespective of the theory's applicability to problems independently raised. It is testimony to the pervasiveness of the thesis that epistemic assessment is reducible to semantics that van Fraassen's conflation of the hitherto quite disparate notions of empirical adequacy and saving the phenomena has gone unnoted.

---

[20] As Duhem, summarizing the state of instrumentalism in 1908, put it: "we now require that . . . [scientific theories] save all the phenomena of the inanimate universe together"—*To Save the Phenomena* (Chicago: University Press, 1969), p. 117.

Much epistemology in our day is arbitrarily and unreasonably constrained by these developments. Our concluding, positive moral is that epistemic warrant unfettered by semantics has rich and varied sources yet to be exploited.

LARRY LAUDAN

University of Hawaii/Manoa

JARRETT LEPLIN

University of North Carolina/Greensboro

PENNY PINCHING AND BACKWARD INDUCTION*

"REASON is the pace, increase of science the way and the benefit of mankind the end." Hobbes makes this remark early in *Leviathan,* meaning it to apply both to the study of human behavior and to human behavior itself. Reason bids each of us do what benefits himself; science helps each to identify the rational choices; the results benefit all. The modern science most deeply committed to this hope is neo-Classical economics, and its most enlightening analysis of reason is supplied by game theory. The general idea is to study economic behavior as choices made by rational agents in allocating scarce resources. In so far as real-world choices are indeed rational, the science is descriptive. In so far as they are not, it is a normative science advising us how to fare better. In either case, the yardsticks are established by exploring an ideal-type world, where all choices are rational choices, in a sense defined by the theory of rational choice and put to work by game theory.

Notoriously, it soon emerges that individually rational choices need not sum to collectively rational results. Here is an example. *A* and *B* play a game where ten pennies (kindly donated by a third party) are put on the table and each in turn takes one penny or two. If one is taken, the turn passes but as soon as two are taken the game stops and any remaining pennies vanish. What will happen, if both players are rational? Offhand one might suppose that they emerge with five pennies each, or with a six-four split because the player with the odd-numbered turns takes two at the end. But game theory seems to say not. Its apparent answer is that the opening player will take two pennies, thus killing the golden goose at the start and leaving both worse off.

The immediate trouble is caused by what has become known as backward induction. Figure 1 sets the game out as a series of choices between moving "across" by taking one penny and moving "down" by taking two (and thus ending the game). The resulting pennies gained by each player are given by the bracketed numbers, with *A*'s put first in each case. Looking ahead, *B* realizes that they will not reach (5,5), because *A* would settle for (6,4). *A* realizes that *B* would therefore settle for (4,5), which makes it rational for *A* to stop at

---

* This paper is part of the Foundations of Rational Choice Theory project in progress at the University of East Anglia with the support of the Economic and Social Research Council (award R 000 23 2269). I would like to thank Robin Cubitt, Shaun Hargreaves Heap, Adam Morton, Judith Mehta, Chris Starmer, and Robert Sugden for their helpful comments on a previous draft.

0022-362X/91/8809/473–88

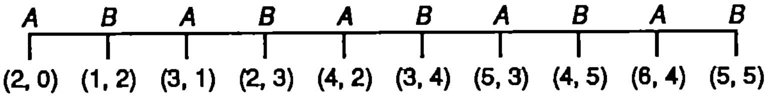


Figure 1

(5,3). In that case, *B* would settle for (3,4); so *A* would therefore settle for (4,2), leading *B* to prefer (2,3); and so on. *A* thus opens by playing down and reason has obstructed the benefit of mankind.

One could try shrugging off this disconcerting result on the grounds that there is nothing odd about it, since, as with the one-shot or finite prisoner's dilemma, rational choices need not be Pareto-efficient in sum. But it is certainly disconcerting. With, say, a hundred pennies on the table, one would swear that rational players would manage to take at least, say, forty-five each before being unnerved by cat-and-mouse considerations, as indeed ordinary people commonly do in experiments with such games. Yet the theoretical proof that (2,0) is the outcome of a game of any finite known length between rational agents is exact and robust, as I shall try to show. At the same time, it threatens catastrophe for both the descriptive and the normative claims of neo-Classical economics, as I shall also seek to show. Yet all is not lost, since there are ways of avoiding the elenchus, if the standard assumptions of game theory are challenged at a finely calculated depth.

I. STANDARD ASSUMPTIONS

Consider a simpler series with only five pennies on the table. The stepping stones of the backward induction are then these:

(1) *B* would take two pennies at his second turn.
(2) *A* would take two pennies at his second turn.
(3) *B* would take two pennies at his first turn.
(4) *A* will take two pennies at his first turn.

To justify and connect them, the assumptions required for a game between ideally rational agents can be conveniently specified thus:

(M) *Motivation*. Each player has complete and consistent preferences for the possible outcomes, ordered on the principle that more for himself is better than less.

(I) *Information*. Each knows the rules and conditions, and can draw all warranted inferences. This includes knowing that both players are motivated as above and are rational in a sense given by the three conditions together.

(K) *Common Knowledge*. Each knows whatever the other knows. (Thus, if *A* knows that *p*, then *A* knows that *B* knows that *A* knows . . . etc. . . . that *p*.)

Given these three assumptions, each stepping stone seems secure and the backward induction is therefore at least tempting.

A preliminary query concerns (M): if the assumptions are meant to specify a notion of rationality, it is not plain that they should make rational agents self-interested in the sense given by (M). For the moment, however, I postpone this topic with the preliminary remark that the players cannot make rational choices without some mutual knowledge of what is motivating each of them. (M) is convenient both because it is a standard assumption and because it helps to spark the backward induction neatly. It does the symptomatic job of giving each player an aim that seems reasonable in itself but oddly self-defeating when the other player has a matching aim.

It may also be asked why I have separated (I), mutual information about the rules of the game, the pay-offs, and the players, from (K), common knowledge. (I) is meant to make the players translucent, so to speak, in ways that are uncontentious, while also raising a question about how far translucency can extend. (K) offers the answer that ideally rational agents are fully transparent to one another, thus clearing the way for the suggestion that the source of the trouble with backward induction lies in the peculiarities of this limiting case.

II. COMMON KNOWLEDGE

Common knowledge, sweepingly defined so that "each knows whatever the other knows," is a condition so strong that it may make it *impossible* for the opening player to take only one penny and hence *impossible* that any later node in figure 1 will be reached. Hence one might protest that it debars *A* from reasoning about what would happen were the game to proceed, since the reasoning would rely on premises that he knows to be false. In that case, the assumptions that generate the paradox would need restating and the question would become whether, as Jon Elster[1] puts it, "there exists a set of assumptions about the players which are strong enough to allow us to infer what the players would do at various nodes of the game while also weak enough to be consistent with those nodes being reached" (*ibid.*, p. 5).[2]

Philip Pettit and Robert Sugden[3] propose a restatement of the common knowledge assumption that disposes of the backward induction paradox. They start by arguing that "under the assumption of common knowledge neither player is allowed to think strategically" and hence that the paradox cannot get started. Translated for the penny game, the nub of their argument is this. Given (K), *A* must open by taking two pennies. For were he to take only one, *B* contemplating this "pass" and applying backward induction, would conclude that *A* had acted irrationally. But *B* ***knows*** that *A* is rational. Hence there is a coherent tale to tell only if *B* starts merely with a ***belief*** that *A* is rational and can be led by an opening pass to revise this belief. Equally, *A* can coherently contemplate later nodes being reached only if he at first ***believes***, but does not ***know***, that *B* is rational, and stands ready to revise that belief, if *B* were to pass. But, when common knowledge is replaced with revisable common belief, so that backward induction has possible outcomes to work on, it ceases to be uniquely rational to open by taking two pennies and the backward induction is thwarted. Therefore, the paradox either does not arise (common knowledge) or is resolved (common belief).

---

[1] *The Cement of Society* (New York: Cambridge, 1989).

[2] Elster sets the problem out crisply and cites an unpublished paper by Christina Biccieri. See also K. Binmore, "Modelling Rational Players," *Economics and Philosophy*, III, 4 (1987): 179–214, for part I; IV, 1 (1988): 9–55, for part II.

[3] "The Backward Induction Paradox," this JOURNAL, LXXXVI, 4 (1989): 169–82. In what follows, I shall take my cue chiefly from their fine paper, which I recommend warmly even while disputing its thesis.

I agree that complete mutual transparency is a highly peculiar condition, even for an ideal-type, limiting case. But the oddity does not lie in the logic of the concept of knowledge. Pettit and Sugden claim that, under the assumption of common knowledge, both players in a one-shot prisoner's dilemma "***must*** defect" (*ibid.*, p. 181). 'Must' is too strong. Even if they ***will*** defect, given common knowledge, it does not follow that they ***must***. "Necessarily, if *S* knows that *p*, then *p*" does not imply "If *S* knows that *p*, then necessarily *p*." When gods play chess, each knows that the other ***will*** make no mistakes and, in this knowledge, identifies inferior moves by considering what would certainly happen if they were made. This ***reductio ad absurdum*** reasoning proceeds by contemplating positions that will not be reached. It could not proceed if such positions were impossible. But they do not become impossible as a consequence of the players' being omniscient.

The peculiarity of common knowledge seems to me to lie not in its logic but in the sheer transparency that it confers on the players. Any reason that might move *A* to pass would be known to *B* and vice versa. Any traps and strategems would be instantly detected. This would certainly take the steam out of zero-sum games like chess. But even common belief does the same for simpler ones like tic-tac-toe, where humanly rational players always draw. The penny game, however, like the prisoner's dilemma, is not zero-sum. Both players can do better than play safe if and only if ***both*** players will rationally risk it. Pettit and Sugden are, I think, right to say that it can be rational to contribute to the better outcome only if the players are somewhat opaque to one another. But it is not plain that the doubt thus licensed operates by shaking an initial belief that the other player is rational.

The general idea is that an opening pass by *A* would inject an indeterminacy whose upshot is that *A* no longer has a uniquely rational opening move. I can see three ways to make the case but find none of them convincing. Firstly, an opening pass might lead *B* to judge *A* utterly irrational, thus withdrawing a premise needed to run the backward induction and so letting *B* pass in the hope of a further pass by *A*. Taken in isolation, however, this dissolves all scope for reasoning ahead, since *B* has to judge *A* to be extremely stupid. Otherwise, since the players are still motivated (M) to prefer more to less at the end, *A* would still opt for (6-4) rather than (5-5) and *B* would still pre-empt this. Unless *A* is demented, this is enough to activate the backward induction.

Secondly, then, an opening pass might lead *B* to credit *A* with rationally permissible beliefs that would warrant a pass. These beliefs must include a belief that *B* will pass in response but can vary in the reasons that *A* attributes to *B*. Thus, *A* might expect *B* to judge *A* demented, as above, or might expect *B* to conjecture that *A* is following one of various rules, each of which involves at least one further pass by *A* (although differing in how many). Provided that *B* has several possible conclusions and is unsure whether or how completely *A* is rational, then this prospective uncertainty can enter *A*'s original calculation. Unfortunately, however, the previous snag stands. If *A*'s plan is well laid, *B* puts a stop to it by killing the game one move before *A* intends, with backward induction ensuring that this means at once. *B* does the same, if *A* is irrational, as above. Hence the second option leads nowhere.

Thirdly and most plausibly, *A*'s opening pass might show him to hold rationally permissible beliefs that neither forbid a pass nor require one. They must include a belief that *B* may (not will) pass in response. This belief by *A* is warranted because indeterminacy breeds indeterminacy not only in *B*'s mind but also, through feedback, in the mind of *A* doing the original calculation. This amplifies the curious element of self-reference in the expectations of each player to the point where the backward induction loses its grip and each is warranted in whatever he does. Paradox is met with paradox, thus inviting analogies with, for instance, the surprise examination, where the pupils' *reductio* of the teacher's promise seems to reinstate the promise. That sounds like the right sort of place to locate the trouble.

But the cure seems to me too drastic, if the price of stopping the backward induction is to leave rational agents without a rational course of action. Game theory might do better to hang on to the disease! Consider the closing stages of the penny game, if indeterminacy is injected into the opening stages. Would *A* still take two pennies, when the only remaining outcomes are (6, 4) and (5, 5)? Would *B* still pre-empt him at the turn before, so that *A* would still cash in at the turn before that? Pettit and Sugden presumably mean indeterminacy to reign here, too. In that case, consider the very simple game (of the penny sort but with slightly different rules and pay-offs) in figure 2. Could *A* pass? The standard answer is that he could but, being rational, will not. If *A* were to pass, would *B* have a determinate response? The standard answer is that *B* would play down, because he prefers (0,3) to (2,2). To reject these answers is to destroy the roots of game theory.

Meanwhile, there is a less destructive line to pursue, and I adjourn further discussion of reason until more has been said about (M). That more pennies are better than less has, so far, prompted *A* in figure 1 to settle for (2,0), because he prefers (6,4) to (5,5), thus somehow losing sight of the fact that (4,5) would have been better for him than (2,0). Can we restore his common sense by a challenge to the standard game-theory idea of what motivates rational players?

III. MOTIVATION

Game theory is a formal system in which a "game" is a set of choices by instrumentally rational agents, where the pay-off to each depends on the choice made by others. This system seems to have quirks, as when rational choices sum to outcomes inferior for all. The quirks are teasing because such "games" invite thought about the games of flesh-and-blood social life. Thus, the penny game between *A* and *B* illuminates the self-defeating character of some everyday encounters between, one might say, Jack and Jill. At the same time, however, one notices that Jack and Jill do not always land themselves with the disappointments of *A* and *B*. Is the difference simply that *A* and *B* are more rational than Jack and Jill, or has the style of abstraction filtered out something crucial to social interaction? With this question in mind, I turn to motivation.

In the analysis of the penny game, motivation is conceived in Humean terms. Reason alone cannot be a motive to any action of the will; it is the slave of the passions and its only office is to serve and obey them. Whether *A* and *B* choose rationally depends not on

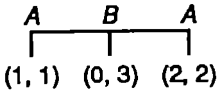


Figure 2

what their passions are (provided that they are consistent) but on how well their choices satisfy their passions. To perform the calculations needed, however, there has to be a unifying measure of satisfaction, standardly called "utility." Although conceived originally as a psychic flow or inner glow, utility now merely marks the presence of a preference; but the claim to unity is preserved by specifying complete and consistent preferences in a rational agent. By keeping silent about sources of utility, moreover, a universal analysis of motivation can be offered, applying equally to egoists and altruists, patriots and cowards, Christians and atheists. Every rational agent is then wholly self-interested, in the very precise sense that he endeavors to maximize his utility.

This means that *A* and *B* play the penny game not for pennies but for what we can call, with seemingly harmless poetic license, utils. Jack and Jill, eyeing figure 1 and thinking in pennies, might not care about one less or more or might actively prefer a fair outcome of five each. *A* and *B* have no such room for maneuver. For them figure 1 gives pay-offs in utils, which are determined in advance of play by taking any and every source of satisfaction into account. The numbers shown represent the given preferences of *A* and *B*, scored in utils, over the outcomes in pennies. In the present case, the utils and pennies happen to match. Were it not so, then the utility numbers would be the ones that mattered.

Utility numbers thus have a dual function. They serve as information available to both players and they also motivate rational choices. The motivating function is, so far at least, wholly automatic. If the game ever reached the penultimate node, *A* would automatically take two pennies, thus securing himself a (6,4) outcome. This, I submit, is all it takes to animate the backward induction. One might have hoped that reflection could rescue them by pointing out the mutual benefits of an indirect strategy, whereby each distances himself from the utility numbers and acts as if upon a different set of figures. But there is simply no scope, while the numbers have their dual function.

That may seem too quick. It is certainly too quick for the godfather of this analysis of motivation. Hume remarks in his essay, "Of the Original Contract":[4]

> For, as it is evident that every man loves himself better than every other person, he is naturally impelled to extend his acquisitions as much as possible; and nothing can restrain him in this propensity but reflection

---

[4] In E. Barker, ed., *Social Contract: Essays by Locke, Hume, and Rousseau* (New York: Oxford, 1956).

> and experience, by which he learns the pernicious effects of that licence, and the total dissolution of society which must ensue from it (*ibid.*, p. 277).

He credits "reflection and experience" with the power to prevent otherwise pernicious effects, helped along by the fact that self-love is not our sole motivating sentiment. Adam Smith works the same street in *The Theory of Moral Sentiments*,[5] suggesting at one point that we have something of an impartial spectator in us—enough to secure "the harmony of society":

> However selfish mankind may be supposed, there are evidently some principles in his nature which interest him in the fortunes of others . . . . By imagination we place ourselves in his situation (*ibid.*, p. 9).

But, once utility numbers have been assigned, the sentiments that occasioned them have no further part to play. If reason's only task is to point the agent toward the larger number, then reflection and experience are powerless and the impartial spectator can only weep.

This may still seem too quick. Even if some sentiments, like the natural sympathy that Hume thought a crucial element in human nature, affect only the assignment of utilities to outcomes, others, like a willingness to view ourselves impartially, might intervene between utilities and actions. After all, the utilities in the penny game are not *expected* utilities, as is all too plain if the expectation of reaching the later nodes is nil! So can "reflection and experience" not stop a rational agent applying the principle that more for himself is better than less so as to emerge with less? Reflection notices that it would be rational for *A* to start by passing if and only if he expected *B*, moved by parallel reflection, to pass. In many similar games, there are several sets of reciprocal expectations that offer the players a grip on each other's bootstraps. In the penny game, is there not enough of an interlock of expectations to let reflection coax *A* and *B* past the first obstacle?

I fear not. So far, *A* is so motivated that he can be relied on to take two pennies if only two remain, and *B* is so motivated as to deny him the chance. An initial pass by *A* could affect these reliable expectations only if it could count as a credible commitment or threat. But, since there is no scope for reprisals and no way for a player to burn his boats and shut off the troublesome endplay, it remains foolish to cooperate. Credibility implies that both players can rationally rely on the game ending with (5,5) as a result of rational choices; and neither *A* nor, therefore, *B* has any reason to expect this amiable outcome.

---

[5] New York: Oxford, 1976.

That narrow-minded retort, however, depends on a further feature of the motivational assumption. Implicitly the scene has been so set that only forward-looking reasons can motivate. Commitments, and indeed threats, involve backward-looking reasons. If Jack makes Jill a promise, he undertakes to let the fact of having promised move him at a later date, even though a forward calculation at that point might show it more profitable for him not to. To threaten Jill credibly, he must convince her that her having done what he had warned against doing will prompt him to retaliate even at a loss. None of this apparatus is so far available to *A* and *B*. What *A* can expect *B* to do, and vice versa, depends solely on what will lie ahead at the point of decision, as if each turn were the start of a new game. Even if *A* is allowed to utter the words of a promise or threat, the words change nothing, because *A* will do what he would have done without them, namely, make the most rational move and let bygones be bygones.

The forward-looking character of reasons in game theory is less obtrusive for repeated games. Where there is more of a future to play for, it may make sense to invest in a reputation for constancy even at some immediate cost. But the reasons for doing so are still forward-looking and they carry conviction only by reference to improved prospects. Even in supergames no one does what he has promised solely for the reason that he has promised, and, in so far as the concept of promising requires that having promised can be a sole reason, supergame practices are borrowing from a world beyond game theory. Similarly, the "ethical" preferences sometimes invoked to account for otherwise irrational choices can work only by affecting the utility numbers before the game starts. In the analysis provided so far, they are simply an element in a preference set that is complete, consistent, and a matter of public information. There is no scope for giving them special influence during play, once the utility of outcomes has been established in advance.

A Rawlsian veil of ignorance sounds promising here. If *A* and *B* know everything about the game, its pay-off, and the forward-looking reasons that they will activate, but do not know who will have first move, then they might see things through the eyes of an ideal spectator. It is as if they were to play the game with *A* starting but then to allot the pay-offs afterward by tossing for who got which. Would they not rationally agree that cooperation leading to (5,5) is the best insurance policy and hence that whoever turns out to be "first player" will pass? Yes, they have forward-looking reasons to assent. But ignorance has not turned egoists into Kantians or even utilitarians, and as soon as the veil is lifted there will be forward-looking reasons not to comply.

The general snag is that it simply does not matter how exactly the moral psychology of the agents works, while utility numbers are given in advance, preferences are complete and consistent, and these two sides of a single coin carry the players irresistably forward. While the utilities have their dual function, otherwise promising thoughts about other sorts of reasons and two-tier moral psychologies where reflection can make a difference are frustrated. Yet, to break their grip, we must stand further back.

IV. REASONS AND MOTIVES

I conclude so far that *A* and *B* are stuck with the backward induction paradox, which needs only fixed pay-offs and an assumption that motivation is supplied by forward-looking reasons addressed to the pay-offs. Yet I do not draw this conclusion for Jack and Jill and the human encounters that game theory idealizes. Try to think of a game of social life that is played by people who know as much and as little about one another as *A* and *B* do. The international stock exchange? It could not function without trust and, hence, even if one is duly cynical, without screening devices to ensure participants who take normative expectations seriously when deciding what to do next. Freak games between complete strangers, as when a lifeboat after a shipwreck has only a tiny stock of water? Disasters do not create a social vacuum large enough to destroy all normative expectations. How about games across the garden fence, where neighborly bonds are only skin deep, as when you and I jointly cultivate a superb melon in the uneasy knowledge that, if it disappears one dark night, the other can get away with blaming vandals? We would never have started, had either suspected the other of being a game theorist!

The basis of economics is often said to be exchange—an interesting idea, if one warms to George Simmel's dictum that "all contacts among men rest on the schema of giving and returning the equivalence." In an exchange, according to game theory, each party performs an action that is rational and the exchange is the sum of these separate choices. The penny game looks at first like an exchange made in a series of turns, each of which is a rational choice. The backward induction puts pressure on what is supposed to connect the series. When Jack and Jill play the penny game, Jack opens by passing, thus making Jill an offer, which Jill accepts by passing, too, and Jack honors his obligation by passing again. How foolish of them, if reasons operate only forward! Yet, in that case, they do not operate at all and Jack becomes *A*, who is led by backward induction to turn down the chance of more.

The contrast between kinds of reason may be clearer with a change of example. Suppose that Aunt Maud passes a hat around to her nephews Tom, Dick, and Harry, promising to double whatever money they put in and then to share the proceeds equally among them. If Tom is a well-tutored game theorist, he analyzes the conditions as a game between *A*, *B*, and *C*, where each invests x pennies and gets back two-thirds of his investment plus two-thirds of everyone else's. Thinking as *A*, he notes that he gets two-thirds of the rest whether he invests anything or not, and makes a loss on anything invested by himself; so he puts nothing into the hat. Perhaps he stops to ask whether a contribution might stimulate one by *B*. But he then reflects that *B* will reason as he has just done, regardless of what is already in the hat, and so will *C*. Aunt Maud can be confident that her amiable gesture will cost her nothing.

If Tom is not so well-tutored, however, fraternal relations should produce a different outcome. Admittedly, it is not plain how much he is to contribute. What if he has 50 pence in his piggy bank, Dick £1 and Harry 10 pence? What if he is the oldest? But, even if the notion of fair play among brothers is ambiguous, contributions of, say, 50 + 75 + 10 will produce 90 pence each and so a profit for all. More simply, if Aunt Maud puts £3 in the hat before passing it round, saying that each may take £1, if he wishes, and that she will double whatever is not taken and share it out equally, then, whereas *A*, *B*, and *C* would each make £1, Tom, Dick, and Harry should each make £2.

'Should' occurs twice in the last paragraph and is tricky. One might construe "fraternal relations" as meaning that Tom, Dick, and Harry are morally or socially bound to help produce the largest collective benefit. But that would take us too far afield. I want to propose the weakest reading of the scene which includes kinds of reasons missing from the atomistic game-theory reading. The crux lies in the difference between normative and rational expectations. So far we have been dealing with the latter. When the players are *A*, *B*, and *C*, what each expects is what each predicts. Even if *A* can somehow borrow from the stock of social practices to signal for an exchange, he can only predict that *B* and *C* will find no forward-looking reason to respond. Bygones are always bygones. Tom, Dick, and Harry, by contrast, are related by social ties governing what they can expect *of* one another. These expectations are normative in the sense that failure to comply attracts moral criticism, especially from those entitled to expect compliance. A moral vocabulary comes into play, even if one relative to a particular social and cultural setting rather than grounded in pure ethics.

In this vocabulary, exchange involves reciprocal obligation, set within an existing scheme of normative relations. By accepting what is offered, the second party acquires a reason to do his bit which it will take a special justification to ignore. Wittgensteinian thoughts arise about the character of action in relation to rules and suggest a fresh sense in which social interaction consists of "games." Although I cannot pursue them here, I mention Wittgenstein as one way of showing how past actions can give rise to backward-looking reasons. If Dick *owes* Tom £5, then this is a reason for him to give Tom £5 which is distinct from any reason having to do with taking pity on Tom, cheering him up, or making a shrewd investment. Tom expects to receive £5 in two different senses of 'expect'. He expects it of Dick as a matter of entitlement and he predicts that it will be forthcoming. That these are different senses is clear, if one notes that Tom can be owed £5 but predict sadly that Dick is not going to pay up. I mention Wittgenstein to indicate that it takes a profound analysis to spell out these homely points.

We should not regard social actors as clockwork toys in the domain of normative relations. There is no need to construe *homo sociologicus* as a throughput between social structure and social action. The playing of social games takes judgment even when one seeks only to play the game aright. We can credit social actors with the power to step back reflectively within the game of the moment and also to step out of it. Tom, Dick, and Harry illustrate both possibilities. The fact that they are brothers signifies only indefinite fraternal expectations, and the commercial character of Aunt Maud's offer makes them wonder how far brotherhood is relevant. Such ambiguities leave plenty of room for reflection and experience on the part of what are nonetheless social actors. The force of calling them "social," however, remains their openness to backward-looking reasons for action. They have the conceptual makings of an exchange for mutual benefit, even though that does not guarantee success in staging one.

Exchange, like many social or moral transactions, requires trust—a reliable normative expectation that today's forward-looking reasons will become tomorrow's backward-looking reasons. Game theorists and contractarians see this as clearly as anyone. So they are tempted to try building on the thought that there are forward-looking reasons to introduce a scheme of backward-looking reasons, rather as rule-utilitarians try stopping the rot that act-utilitarianism spreads in institutions, by the two-tier device of entrenching the most useful rules. It is as if reflection could lead *homo economicus* to beget *homo sociologicus,* who, having internalized the normative expectations surrounding a transaction, reworks the utility numbers before handing back to *homo economicus* to make the choices.

If indirect solutions of this kind worked, they would have the interesting result of shifting the basis of economics from decision theory to the theory of organizations and institutions. Economic agents would enter markets essentially as members of firms, banks, government agencies, unions, churches, and households, since the serious work would be done in the assigning of utilities. But, first, there is still a problem in seeing how an indirect solution could work.

If one asks exactly why backward induction holds in the penny game, the immediate answer is that *A* and *B* are too cramped by the setting to be able to strike a bargain. Even if *A* signals for one by passing, *B* will ensure against the bargain's not being kept or, more simply, will stand ready to exploit it. This suggests that they might get on better if they were less cramped, for instance, if they could talk the game over or make a bargain in front of witnesses. But such human possibilities require a fresh moral psychology, if the problem of compliance (of giving *A* sufficient reason not to stop the game at (6,4)) is not to drive out trust. It has emerged, moreover, that nothing can be done in this direction unless a rational agent can come to override the apparent *diktat* of the utility numbers. In the absence of a moral psychology that allows all this without merely destroying the basis of game theory, the cramped setting of the penny game is fair comment.

So let me now try to answer Elster's question whether "there exists a set of assumptions about the players which are strong enough to allow us to infer what the players would do at various nodes of the game while also weak enough to be consistent with those nodes being reached." I suggest that we start by stating the pay-offs in pennies after all, so as to be able to break the stranglehold of the utility numbers. Then let us make the ranking of an outcome in a player's order of preference depend not only on how many pennies it brings but also on how it is reached. Thus, six pennies gained by breach of trust need not outrank five gained honorably, or indeed the four that *A* gets by acting decently despite *B*'s intended failure to act likewise. This is no simple amendment, however, since it involves importing a normative vocabulary into the description of the outcomes and also refashioning *A* and *B* so that normative descriptions can make a difference.

Condition (M) therefore needs restating carefully. A first thought is that the players can still have complete and consistent preferences over outcomes in advance of play, if the preferences range over routes as well as destinations. But, if we construe this by analogy with a journey where it is worth something to travel in comfort through pleasant scenery, we shall still be dealing in utilities. That, admittedly, keeps the moment of decision clear. In the penny game, for instance, the more profitable outcomes can only be reached by routes where the players make and keep what are in effect bargains. A shorter version played with a total of six pennies and two players who prefer to play fair might look as in figure 3:

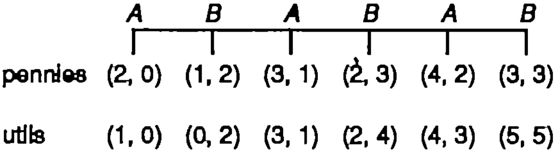


Figure 3

Here a forward induction will keep *A* and *B* going to the end. *A* opens with a pass because he knows that *B* will pass because *A* will pass again because *B* will pass again because *A* prefers (5,5) to (4,3), when the pay-offs are stated in utils.

But this is no solution at all. It merely rewrites the penny game arbitrarily so that it no longer illustrates the problem. That the translation into utils is arbitrary is plain, if one considers the intermediate outcomes, where more pennies dishonorably gained have been ranked in comparison with less pennies honorably gained. Meanwhile, nothing has been done to introduce the backward-looking reasons needed for honorable conduct. We need a fresh version of (M) that leaves the players a possible escape when faced with a threat of backward induction.

A very different version might make rational players Kantians, who choose whatever Adam Smith's impartial spectator would recommend, regardless of their individual profit. Thus, to revert to the game in figure 2, *A* plays across and *B*, being a Kantian, too, also plays across, and they emerge both rich and virtuous. But this seems altogether too easy, not least because the backward induction problem has again simply vanished. Kantians always have a dominant strategy of heading for the right outcome, regardless of what other people will do. Two Kantians will therefore achieve it, but there is no warrant for saddling game theory with a motivational assumption that ensures that only Kantians can play with success.

I introduced figure 2 earlier to make the point that radical indeterminacy would destroy the basis of game theory. If *A* could leave *B* without a rational course of action by passing, then the theory of rational choice collapses. But at that stage motivating reasons were only forward-looking and it made no difference how the players came to arrive at figure 2. Similarly, if *A* were to play across, that would create a new game consisting of figure 2 with the left-hand node deleted. What we now need, I suggest, is a version of (M) where it can matter what happened earlier in a game, because the players can have reason to care, but where compliance with bargains is not guaranteed. This suggests rewriting (M) thus:

> (M*) *Motivation:* each player ranks the possible outcomes by the measure of what best serves the interests of both players, provided that he expects the other player to rank them in the same way.

This version lets *A* open the penny game with a pass, because that will encourage *B* to expect a further pass. But the proviso in it is meant to make the players translucent rather than transparent. As with Pettit and Sugden, therefore, there is an unsettling element of mutual self-reference. But they located it in the initial common belief that the other player is rational, where it spreads like dry rot. By locating it in an initial common belief about motivation, I hope to keep it within bounds. For, although the players again have no certain course of action, they are simply short of information which, if they had, would not introduce a further paradox.

This translucency is due in part to the game-theory abstraction that turns Jack and Jill into *A* and *B*, thus shelling them out of their normal and normative social relations. Since *A* and *B* are strangers, *A* has excellent reason to settle for (1,1) in figure 2 and some reason not to risk even one pass in the original penny game. But Jack and Jill are also translucent, even if they know more about each other. They are *persons* and thus inherently better placed to act in the light of "reflection and experience," as Hume recommends, but still separate persons, whose honorable character is not guaranteed. That leaves much more to be said about the relation of rational agents, as abstracted by game theory, to human flesh and blood than I can attempt here, but I can at least offer a concluding view of backward induction.

V. CONCLUSION

Rational agents, as initially defined, are indeed trapped by backward induction, in my view. Common knowledge is indeed an example of extrapolating to a limiting case that turns out not to make sense. But the trouble does not lie in the logic of the concept of knowledge, since even the omniscient can perform a ***reductio ad absurdum*** and the trouble is still there, if common belief is specified instead. The obstacle to rational cooperation lies in utility numbers like those got in the penny game by letting each player be motivated solely by his own gains. This is a natural approach, given a Humean view of what makes us tick, and a tempting one, granted that we have other sentiments besides greed. But once the numbers are in place and known to both players, and once they are treated as irresistible magnets, there is no escape.

Alternatively, however, Jack and Jill can have backward-looking reasons for action, alien to A and B. These reasons are not sentiments but normative features of a situation that they have partly inherited and partly created. Jack and Jill are *persons* and thus open to reflection and experience, because they are related in ways also alien to A and B. That clears the line for a long and complex comparison of Kantian with Wittgensteinian moral psychology, in the hope of deciding which has the truer account of rational agency.[6]

Game theory, meanwhile, retains its fascination but becomes more provisional. It applies in full wherever pay-offs in goods can be treated almost as if they were in utils; and much can be understood about social interaction through this style of analysis. But it cannot dispel the paradoxes that arise by its own resources. Economics is finally no more a self-contained discipline than the economy is an isolable realm of social life.

MARTIN HOLLIS

University of East Anglia

---

[6] I have tried to set the stage for this comparison in "Moves and Motives in the Games We Play," *Analysis*, L (1990): 49–62; and in "Honour Among Thieves," *Proceedings of the British Academy* (1990): 163–80. Further discussion of moral psychology can be found in chs. 5–6 of *The Cunning of Reason* (New York: Cambridge, 1988).

BOOK REVIEWS

Aristotle's First Principles. T. H. IRWIN. New York: Oxford University Press, 1988. xviii, 702 p. Cloth $75.00.

T. H. Irwin's book presents itself as an attempt to show that Aristotle was, in a certain sense, a systematic philosopher. This view of Aristotle has been widely rejected in recent scholarship for two reasons: first, the naïve assumption of ancient and medieval commentators, that Aristotle's philosophy was based upon a fixed set of fundamental principles in such a way that every part of it must be consistent with every other and had to be understood in terms of those basic principles, seems untenable in the face of numerous conflicts between the doctrines of different treatises. Especially since developmental studies became popular from the 19th century on, it seemed much more plausible to explain such inconsistencies as signs of change or development over the nearly forty years of Aristotle's career as a philosophical writer. Secondly, Aristotle's own remarks (in the *Posterior Analytics*) about the separateness of different branches of science explicitly discourage the assumption that his treatises might be unified by a common systematic background. Irwin does not wish to return to the old naïve view, of course, but he argues that with the appearance of "first philosophy" in the *Metaphysics* Aristotle develops, in contrast with his earlier methodology, a set of metaphysical "first principles" that underlie many of his later specialized treatises. The title of Irwin's book, then, should be understood in two ways. First we get an examination of Aristotle's (early) views on first principles (Part I), then an exposition of the doctrines from the *Metaphysics* which serve as first principles in the later works (Part II). The third part supports the other two by case studies meant to show how Aristotle's metaphysical doctrines are at the basis of his theories of the soul, of ethics, and of politics.

Much of the interest of the book lies in this third part, which offers detailed analyses of the main arguments of the *Ethics* and the *Politics*. One cannot quite help thinking that the epistemological and methodological theme was to some extent superimposed on these studies. And that is just as well, for if one finds the arguments of the first two parts less than convincing, the third will still be worth studying for its incisive and rigorous analyses of crucial doctrines.

Since we have virtually no information about the relative chronology of Aristotle's works independent of their contents, an interpretation that argues for development and changes of mind must be

0022-362X/91/8809/489–96

cautious not to beg any questions by assuming the chronology that best fits the interpretation. Irwin is well aware of this point, and his chronological assumptions are properly modest. He supposes, with most modern scholars, that the treatises traditionally grouped together as the *Organon* are earlier than most of the other extant works. Other treatises are then considered as earlier or later depending on their closeness or otherwise to the philosophical positions of the *Organon*. Here disagreement about interpretation will obviously invalidate chronological conclusions but, as Irwin rightly points out toward the end (472), most of his argument is in fact independent of chronological considerations. Development, as I think most scholars would now agree, offers the most plausible explanation of the difference between the *Categories* and *Posterior Analytics,* on the one hand, and the *Metaphysics,* on the other, and this seems to be the only crucial assumption.

The main argument of parts I and II is roughly as follows. In the early treatises, Aristotle puts forth a set of epistemological and metaphysical views which leads to serious difficulties about the status of his own philosophical doctrines. On the one hand, Aristotle assumes metaphysical realism and epistemological foundationalism; on the other hand, his account of the philosophical investigation of first principles, i.e., basic assumptions of scientific knowledge, seems to show that all he can attain by his method of "dialectic" is a theory that achieves coherence among "common beliefs" (*endoxa*),[1] which will not support a claim to objective truth. In the *Posterior Analytics,* Aristotle offers an intuitionist account of the knowledge of first principles; but apart from being very sketchily stated, this theory is open to fatal objections. Aristotle seems to have no defense against the charge that he cannot distinguish between the correct grasp of a first principle and the mere belief that one has grasped a first principle. His method might suggest that principles must be grasped by inference to the best explanation, but this is incompatible with Aristotle's own assertion that knowledge of first principles cannot be based on inference. Hence Aristotle's own theories of substance, essence, change, and so on seem to rest on no more than what Irwin calls "pure dialectic," an argument that is based on more or less widely held beliefs. Philosophers who do not give much weight to common opinions could therefore simply reject Aristotle's theories as irrelevant. But it seems clear that Aristotle himself presents his views as objective truths, not just common opinion.

---

[1] I would prefer to follow Jonathan Barnes in translating this term as "reputable beliefs," since the views considered include those of other philosophers which may be far removed from common opinions. According to Aristotle, it is the orator, not the dialectician, who has to limit himself to widely accepted beliefs, given the audience he is trying to persuade. But see Irwin, 494, n. 42.

Since intuitionism does not seem to offer a way out of the dilemma, Aristotle revises his methodology and, beginning in the *Metaphysics,* introduces what Irwin calls "strong dialectic," a method that builds upon a subset of common beliefs and relies only on those which an opponent could not seriously refuse to accept. The paradigm case of an argument to show that certain propositions are such that one cannot seriously deny them is Aristotle's defense of the principle of noncontradiction in *Met.* iv. 4. Aristotle argues that anyone who claims to reject the principle will find himself assuming its truth in the very attempt to state its denial. Irwin connects this and the subsequent antiskeptical arguments in *Met.* iv. 4–6 with the *Metaphysics*' project of a most general science of being, claiming that they together establish that anyone who wishes to make a serious claim to knowledge will have to assume the existence of persistent subjects with essences. These, then, will be the proper objects of first philosophy, and in *Met.* vii–ix Aristotle offers a revised theory of substance, built upon more solid foundations than the "pure dialectic" of the *Categories.* The new metaphysical doctrine comes to form the foundation of later treatises, such as the *De Anima,* which argues that the soul is the form and substance of the body, and the "political science" expounded in the *Ethics* and the *Politics,* which support their doctrine of the human good by reference to Aristotle's metaphysical views on substances and natural teleology.

While I would agree with Irwin about the contrasts between the *Categories* and the *Posterior Analytics* as opposed to the *Metaphysics,* I do not think that the hypothesis of a methodological change of mind, or a transition from "pure" to "strong" dialectic is warranted or needed to explain Aristotle's development. Obviously, a revision of an earlier doctrine can be prompted by many considerations, and the changes between the *Metaphysics* and the *Organon* seem to me to be sufficiently motivated by the need to make room for the new concepts of matter and form, on the one hand, and by a consideration of the status of philosophical theorizing with regard to the special sciences, on the other. In the *Topics*—Aristotle's official handbook of dialectic—Aristotle says that dialectic has a "way towards the principles of all disciplines" (*Top.* 1, 101b3) in virtue of its capacity to examine propositions of all kinds. He does not say, however, as Irwin seems to assume, that it is also dialectic that discovers the first principles by offering a theory that will accommodate or explain the "reputable beliefs" from which an investigation

CALCUTTA
CENTRAL

must begin. Theory building is the task of the philosopher, and Aristotle's theories seem to conform to this model—he accepts a theory if it succeeds in preserving or justifying the most widely held or authoritative beliefs and explaining the rejected ones by showing how they arise from understandable errors. Now, according to Irwin, such coherence will not suffice to satisfy the demands of Aristotle's objectivist epistemology, for there could, of course, be several different theories that meet the requirement of coherence. Furthermore, since not all of the reputable beliefs can be true, it is possible that a proposed theory is coherent, but accepts the wrong subset of common beliefs. But it is not clear to me that Aristotle would have been aware of the first point and, as to the second, one might appeal to Aristotle's epistemological optimism, expressed repeatedly in the statement that the opinions of all reasonable people must contain some grain of truth (see, e.g., EE i.6, 1216b28-31; EN i.8, 1098b27-9). This is why he insists that a good theory should not only preserve as many as possible of the reputable beliefs, but also try to explain those which are rejected, on the ground that, if they are not true, they must at least be understandable misapprehensions of the truth, not mere madness and fantasy. Given this optimistic view, Aristotle's position in attacking Plato or Democritus for their contempt of common beliefs is stronger than Irwin admits. Finally, instead of diagnosing a rejection of the intuitionist account of the grasp of first principles in the *Posterior Analytics,* we might fit this into the methodological picture by supposing that we come to grasp the basic principles of a science by realizing that they explain the beliefs from which we had started. Irwin thinks that Aristotle must reject such a move because he maintains that first principles are grasped "noninferentially," and the fact that a certain set of principles offers the best explanation would clearly constitute a reason for accepting them, rather than mere intuitive insight. But I do not think that Aristotle rules out this kind of reason. When he argues that the first principles of a science must be "indemonstrable" and hence grasped without proof, he merely excludes the possibility of deductive demonstration, and an argument to show that we have good reason to believe that p need not be an argument to show that p is the case, let alone an explanation of why it is the case that p.

So far I see no reason to think that Aristotle was caught in a methodological dilemma. The methodology just described seems to be the same as what we find in later treatises. What is new in the *Metaphysics* is the thesis that there can be a discipline, called "wisdom" or "first philosophy," that is most general and hence prior to all the specialized sciences, and deals with being qua being. The *Posterior Analytics* had denied the possibility of such a super science on the ground that the first principles of each science are independent and not derivable from any other science. In the *Metaphysics,* Aristotle declares that such a science is possible after all, and that it also deals with the so-called "common principles," i.e., propositions like the principle of noncontradiction that are accepted by each of the special sciences. We do not know, since Aristotle does not tell us, why Aristotle changed his mind about the possibility of a most general discipline, but it is a fair guess that he might have come to see that his own ontological doctrines, put forward with great confidence, for example, in the *Categories,* did not fit into the picture outlined in the *Posterior Analytics.* However that may be, it seems clear that the investigation of first principles cannot be the task of a demonstrative science, since every such science must assume its axioms and cannot justify them. So Aristotle has to resort to the kind of "negative demonstration" offered in defense of the principle of noncontradiction in *Met.* iv, or indeed to his old method, dialectical examination of reputable views and construction of a theory that will either justify them or show them to be understandable errors. It seems worth noting that the "negative demonstration" of *Met.* iv. 4 is rather untypical for Aristotle and not (as far as I can see) paralleled elsewhere in the *Metaphysics.* On the other hand, it seems to be inspired by Plato's arguments against Protagoras in the *Theaetetus.* Being, like Plato's arguments, mainly a refutation (first of people who pretend to deny the principle of noncontradiction, then of skeptics about knowledge), the series of arguments in *Met.* iv. 4–6 is certainly not officially presented as establishing a fundamental principle about possible objects of knowledge. (The Kantian association is intended here: see 540, n. 18.)

Hence it seems rather bold to assume that this principle—namely, that there must be persistent subjects with essences—established by "strong dialectic," serves as the basis of Aristotle's subsequent disquisitions about substance in *Met.* vii–ix. As far as the modest results of bk.iv will go, a Platonist, or someone who dislikes multiplying entities beyond necessity, could perfectly well shrug off Aristotle's immensely complicated efforts to find a candidate for the role of substance (basic entity) which will be both a basic subject and an essence. Irwin's succinct account of Aristotle's line of argument is an admirable tour de force (no doubt it will be hotly debated by the numerous experts in this particular field) which seems persuasive to the extent that it shows how Aristotle might have arrived at his view (if it is indeed his considered view) that particular forms are the best candidates for the contested position. But given the large number of strange premises (the most mystifying, to my mind, is that "basic subjects must be identical to their essences") that need to be added to reach this result, it can hardly count as simply derived from the epistemological arguments of *Met.* iv. The discussion of substance in the *Metaphysics* is certainly far more sophisticated than the one in the *Categories,* as it has to be once the distinction between matter and form is introduced, but as far as I can see, its epistemological foundations are the same.

Irwin's second illustration of the gap between pure and strong dialectic seems to me entirely unconvincing. Irwin claims that the first two books of the *Physics* are a good example of the weaknesses of pure dialectic in that Aristotle works only with common beliefs, and hence fails to see the deeper implications of his own doctrine of the four "causes." I think that the doctrine is not as muddled as Irwin makes it out to be, and so it is not surprising that it is found in exactly the same version in the—allegedly more advanced—*Metaphysics.*

According to Irwin, Aristotle involves himself in an inconsistency, succinctly described as follows:

> Aristotle seems to accept this inconsistent set: (1) Form and matter are origins of change. (2) The origin of change is the same as the efficient cause. (3) Therefore form and matter are efficient causes. (4) The formal and material causes are not efficient causes. (5) Therefore form and matter are not efficient causes. (6) Therefore they are not origins of change. To defend (1) he must reject (2) or (4) or both (517f, fn. 1).

But (2) is not Aristotle's view, and hence there is no inconsistency. Irwin thinks that (2) must be Aristotelian because Aristotle describes the matter and form of a natural thing as "origin (or principle—the Greek word is the same) of change," and also the efficient cause (*Phys.* ii. 3, 195a11 = *Met.* v. 2, 1013b11; the standard formula is slightly different). It does not follow, however, that all origins or principles of change are efficient causes. The simplest way to show this is presumably by appeal to *Met,* v. 1, the chapter that lists different uses of the term *arche* ("origin" or "principle"). There Aristotle explicitly distinguishes between internal and external principles (of change, as his examples show). To clinch the point, the external principle is illustrated both with the standard formula that introduces the efficient cause and with the standard examples: parent/child, insult/battle. Form and matter are, of course, internal principles of change (of the thing constituted by them), and hence they are not efficient causes (of the same thing). That does not preclude the form from being the efficient cause in relation to another thing: as Aristotle often remarks, the form *in the parent* is the efficient cause of the offspring; which shows, incidentally, that Irwin's (5) does not follow from (4). In general, it is as mistaken to assume that "form" means the same as "formal cause" as it is to think that "principle of change" means the same as "efficient cause." When Aristotle says (*Phys*. ii. 7, 198a25-27) that in natural things formal, final, and efficient cause often coincide, he means that the form has three different explanatory roles, not that there are really only two kinds of causes. There is therefore no need to attempt a reduction of the four causes to the efficient cause, nor to speculate, as Irwin does, that Aristotle has overlooked the crucial role of the efficient cause because he is paying too much attention to common beliefs. If Irwin were right, Aristotle's doctrine would conspicuously fail to cover his own paradigm case of scientific demonstration through explanatory premises, namely mathematical proof. There is no room for material, final, or efficient causes in mathematics, since mathematics deals with immaterial and unchanging entities; the only sort of "cause" in this field must be formal, as exemplified in such propositions as that a triangle has angles equal to 180° *because* it is a triangle.

There seems then to be no good reason to explain Aristotle's development by appeal to alleged revisions in his methodology. Presumably dialectic was never the only method of philosophy, since it does not include theory building, and the epistemological foundations of the search for first principles may well have remained the same throughout, except for the introduction of "first philosophy."

Now, fortunately, one does not need to agree with the developmental hypothesis to accept the main thesis of Irwin's part III, namely, that Aristotle uses his metaphysical doctrines of substance, nature, teleology, and so on in many of his other works. That he does so seems overwhelmingly evident in the *De Anima*, but it is also true, I think, of the *Nicomachean Ethics* and the *Politics*, where Irwin supports his case by showing how Aristotle's theory of the human good relies on his natural teleology. One might only wish to modify Irwin's claim by saying that the metaphysical principles could come from the *Physics* just as well as from the *Metaphysics*.

Irwin's extended analysis of Aristotle's "political science" (ethics and politics taken together) is penetrating and often illuminating. In the *Ethics*, he pursues the question whether Aristotle can justify his apparent assumption that human virtue, conceived as perfection of rational agency, is identical with the set of commonly recognized moral virtues. In the *Politics,* he shows in several cases that, if one pays attention to the difference between basic principles and empirical assumptions, one can see that some of the rather appalling conclusions do not follow from the principles, and hence provide no reason to reject the entire theory as ideologically biased. I would not agree with Irwin's answer to the first question, but it would take too long to argue this in the space of an already long review, and the reader should look for herself.

What remains of the thesis that Aristotle was a systematic philosopher? If it is sufficient for this claim that Aristotle should be using the same metaphysical framework in many of his works, wherever it is relevant, one might find the thesis rather uncontroversial. But Irwin clearly wishes to establish more—namely, that Aristotle consciously developed his metaphysical doctrines as a foundation for his subsequent theories, and that he himself considered them as the basic framework within which one should evaluate and understand the more specialized theories. (The relation would be similar to that between Kant's *Critique of Pure Reason* and his moral philosophy.) If this were true, then one should expect Aristotle to pay far more attention than he actually does to the implications of his metaphysical premises within the respective special fields. As Irwin admits, Aristotle never even explicitly states the—legitimate and important—question Irwin raises about the relation between virtue as perfection of rational agency and the traditional moral virtues. Hence one suspects that Aristotle himself, though freely taking over propositions from his metaphysics where he could use them, did not see those "first principles" as a kind of theoretical backbone that ought to structure the rest of his discussion. Perhaps he ought to have been a systematic philosopher; but I very much doubt that he tried to be one. Whether this was a fault or a blessing will have to be decided case by case.

As far as Irwin's book is concerned, I think that the parts turn out to be far more interesting than the whole. Irwin's interpretations are always worth reading and thinking about, whether one agrees with him or not, and he is surely right in saying that the sort of discussion he offers is needed if one wants to take Aristotle seriously as a philosopher and not just treat him as a museum piece. No scholar with a serious interest in Aristotle can afford to ignore this book.

GISELA STRIKER

Harvard University

What Reason Demands. RÜDIGER BITTNER. Theodore Talbot, trans. New York: Cambridge University Press, 1989. ix + 198 p. Cloth $39.50, paper $13.95.

This book is a revised and expanded edition of Rüdiger Bittner's *Moralisches Gebot oder Autonomie*.[1] It is divided into seven chapters, each beginning with a concise summary of its contents. The first chapter sharply poses and explains the main issue: Does reason demand that one be moral? The next three chapters criticize contemporary attempts to establish an affirmative answer. The fifth chapter maintains that Kant's principle of autonomy, suitably modified, provides a necessary and sufficient condition for moral laws, but it concludes moral laws are "null and void," "practically irrelevant" (106, 107). The remaining chapters develop and defend the radical thesis that what reason demands is only prudence, not morality.

Bittner begins by arguing that the crucial question for moral philosophy is not "Why should I be moral?" but rather "Should I be moral?" Moral demands are construed as familiar kinds of utterances, like insults and greetings, and so what is in doubt is not whether moral demands exist but whether they are "justified" or "valid." Bittner distinguishes between justifying a demand and demonstrating its validity. To justify a demand is to give reasons *why* one ought to heed it, that is, to say what is good about doing as prescribed. To prove the validity of a demand we need only to give appropriate reasons *that* one should conform, and the sort of reasons that are relevant will depend on the kind of demand in question. A demand that one present one's automobile for inspection, for example, may be justified by the fact that conformity will improve road safety, but the validity of the demand can be shown by citing regulations for vehicles enacted by an authority with jurisdiction.

"Why should I be moral?" asks whether moral demands are justified, and "Should I be moral?" asks whether they are valid. Philosophers have failed to give convincing answers to the first question, according to Bittner, and so it is best to focus on the second. Here Bittner concedes that moral demands could be "valid" even though "unjustifiable." That is, conceivably we could have reasons for thinking *that* we should heed moral demands even though we lack reasons that tell us *why* (in the sense that they explain what is "desirable" about heeding such demands).

Although to show the validity of moral demands one would not

---

[1] Freiburg: Alper, 1983.

0022-362X/91/8809/497–501

need to provide "justifying" reasons, one would still need to state reasons of some kind appropriate for moral demands. Unlike the demand for automobile inspection, moral demands obviously cannot be shown to be valid by citing official regulations. As commonly understood, moral demands implicitly make strong claims that severely constrain what could count as a demonstration of their validity. To show that a certain moral demand is valid for a person, then, it is not enough to point out that the person has adequate reasons *of some kind* to do what is demanded. Merely prudential reasons, for example, would not validate the demand *as moral;* and, on Bittner's view, even if all rational agents would agree on moral rules prescribing the action, that would not validate the prescription *as a demand* on people actually unwilling to comply.

In chapters II–IV, Bittner tries to show that the validity of moral demands cannot be based on essential human attributes or accomplishments, actual or hypothetical contracts, or promises. The argument here proceeds primarily by way of a critical review of the views of contemporary writers, including Kurt Baier, Neil Cooper, Philippa Foot, Alan Gewirth, Geoffrey Russell Grice, R.M. Hare, A.I. Melden, Thomas Nagel, Jan Narveson, Richard Norman, John Rawls, David A.J. Richards, and John Searle. Bittner finds none that provides adequate grounds for accepting the validity of moral demands. His objections are crisply stated and pertinent, though often familiar and too brief to be convincing without more discussion.

The wide-ranging scope of these critical chapters should be evident from the following summary of the main points. Gewirth mistakenly assumes that regarding something good entails claiming a right to it. Norman wrongly infers the validity of moral demands from the claim that actions can be understood only by moral concepts. Nagel confuses the weak conclusion that one can meaningfully extend practical judgments from one's own case to others with the stronger claim that the extended judgments are valid. What individuals in Rawls's original position would agree upon is hard to determine, and in any case agreement among fictitious individuals would give us no reason to accept the corresponding moral demands as valid. Searle's argument from the supposed analyticity of "Promises ought to be kept" has been undermined by counterarguments from Hare and others. Prichard relies on an unintelligible notion that one can put oneself under obligation, or bind oneself irrevocably, by one's own action; but in "promising" one can really only communicate intentions and so, strictly speaking, there are no promises.

Finding no satisfactory arguments in contemporary sources, Bittner turns back to Kant. No philosopher, of course, could disagree with Bittner's final conclusions more than Kant, but, as Bittner recognizes, Kant himself forcefully articulated the challenge to moral philosophy that Bittner believes cannot be met. The crucial elements are contained in the second chapter of the *Groundwork*. First, analysis reveals that in judging that an act is a duty one implies that the act is inescapably required by reason, independently of prudential considerations. Second, such rational requirements are conceived and experienced as internal constraints on one's will to which one cannot be indifferent; and so these requirements cannot be equated with conventional rules, the commands of others, or any natural facts. Third, one would be subject to such constraints if but only if they were self-imposed, that is, willed for oneself by oneself as a rational agent with autonomy. These conceptual points, Kant thought, pose a challenge that previous moral philosophy never adequately faced: What grounds are there for believing, and indeed how could it be, that there are constraints on our choices that are (in the relevant senses) rational, not optional, and yet self-imposed? If the challenge cannot be met, Kant insists, the conclusion must be that ordinary moral judgment rests on a false presupposition and so duty is a mere illusion.

Kant, of course, believed that he gave an adequate response to this challenge in the third chapter of the *Groundwork,* or at least in the first part of the *Critique of Practical Reason*. Kant's arguments may be unconvincing to contemporary moral philosophers, but most manage to avoid the skeptical conclusion that there are no duties by rejecting one or another of Kant's initial presuppositions. Bittner, interestingly, does not reject the basic ideas behind Kant's challenge but instead argues that, suitably interpreted and modified, the problem Kant posed is real, that Kant failed to resolve it, and that indeed it cannot be resolved. Bittner's conclusion, surprisingly, is not that there are no valid moral *laws* but rather that there are no valid moral *demands*. A modified Kantian principle of autonomy states necessary and sufficient conditions for moral laws to be valid for a person, and these conditions may sometimes be met. But moral laws cannot support moral demands (alleged "duties" and "obligations") because they have no rational claim on those who do not want to act on them. For those willing to comply, moral rules impose no *demands;* for those unwilling to comply, they are not *valid*.

Imagining a child tempted not to share candy with his brother and unmoved by preliminary answers, Bittner gives the ultimate response that encapsulates his view: "To the child's question 'Why should I be moral?' our answer was this: No reasons can be seen. To the new question, 'Should I be moral?', we now have the answer: If you want to be, yes; otherwise, no" (107).

For those willing to accept enough of the Kantian background to acknowledge prima facie force in Kant's challenge, the crucial steps in Bittner's argument will be his interpretation of autonomy and his modified Kantian autonomy principle. Autonomy, on Bittner's view, is not a duty or an ideal; it is simply a property of agency manifested whenever we act intentionally, according to our ideas and plans. His principle of autonomy states that moral rules are valid moral laws for a person if and only if both (1) all rational beings could agree to the rules and (2) *the person in question actually adopts the rules as guides for his own conduct.*

The last clause, which is the key to Bittner's unKantian conclusions, is supposed to be a refinement of Kant's idea that moral laws are "self-imposed." Kant clearly did not intend the autonomy condition as Bittner understands it, for Kant did not suppose that one could render a moral rule invalid in his case simply by wanting and choosing to disregard it. Bittner realizes, of course, that his autonomy principle is a modification of Kant. What he offers in support of this crucial principle is primarily a critique of Kant. The argument, in effect, proceeds by first assuming an initial plausibility in the Kantian idea that autonomy is necessary for moral laws, next interpreting and criticizing Kant's attempts to articulate the appropriate autonomy condition, and then accepting the residue (however implausible) as true. The argument may be challenged at each stage, but especially dubious, I think, is the move from "self-imposed laws" to whatever fits with an agent's intentions. If Kant's metaphor of self-legislation has any initial plausibility at all, it does so only under an interpretation quite incompatible with the idea that one "legislates" to oneself whatever one wants. Any interpretation, or modification, of Kant's idea of autonomy that reduces self-legislation this far undercuts the initial grounds for supposing that autonomy is a necessary condition for moral validity.

In a brief review one can do little more than explain an author's project, main conclusions, and strategy of argument, but *What Reason Demands,* in my opinion, merits more sustained critical attention. Bittner's discussions of Kant and others are informed and challenging even when one disagrees, and he presents the case for his own radical thesis forcefully and concisely. The last two chapters of his book are constructive and rich with fresh ideas on prudence, developing the theme that "practical reason is not a judge, but a physician" (128). In all, despite substantial disagreements, I think that moral philosophers ought to read Bittner's book. That is, reason advises this, even if it cannot demand it.

THOMAS E. HILL, JR.

University of North Carolina/Chapel Hill

NOTES AND NEWS

The editors regret to announce the untimely death of Alan H. Donagan, the Doris and Henry Dreyfuss Professor of Philosophy at the California Institute of Technology, on Wednesday, May 29. Professor Donagan joined the faculty at Caltech in 1984, after spending 14 years at the University of Chicago, where he was the Phyllis Fay Horton Professor of Humanities. Previously he was a professor at the University of Illinois/Urbana-Champaign, Indiana University, and the University of Minnesota. At both Indiana and Minnesota, he served as chairman of the philosophy department. Earlier, he held academic appointments in philosophy at University College in Canberra, Australia, and at the University of Western Australia in Perth.

Behavior and Philosophy, a journal for the interdisciplinary philosophical study of psychology and behavioral science published by the Cambridge Center for Behavioral Studies, is pleased to announce that Max Hocutt will succeed George Graham as editor, commencing with Volume XX, 1992. All correspondence concerning the journal, including manuscripts for submission, may be sent to Max Hocutt, Phil. Dept., Univ. of Alabama, University, AL 35487.

Stanford Humanities Center will offer up to six external fellowships for 1992–93 intended for postdoctoral scholars and teachers in the humanities, or those in other fields working on related projects, who are interested in spending the academic year at Stanford. The fellowships are primarily intended to enable fellows to pursue their own research and writing; however, recipients are also expected to devote about one-sixth of their time to teaching or in some other way contributing to intellectual life at Stanford. The deadline for application is November 15. Application materials and further information may be obtained by writing to Fellowship Program, Stanford Humanities Center, Mariposa House, Stanford University, Stanford, CA 94305-8630.

NEW BOOKS

ADAMS, E. M.: *The Metaphysics of Self and World: Toward a Humanistic Philosophy*. Philadelphia: Temple, 1991. xv, 325 p. Cloth $54.95.

AVNI, ORA: *The Resistance of Reference: Linguistics, Philosophy, and the Literary Text*. Baltimore: Johns Hopkins, 1990. xv, 313 p. Cloth $39.95.

BROADIE, ALEXANDER: *The Tradition of Scottish Philosophy: A New Perspective on the Enlightenment*. Savage, MD: Barnes & Noble, 1990. vii, 144 p. Cloth $42.00.

CARMIE, STEWART ZANE: *Vulgarity and Authenticity: Dimensions of Otherness in the World of Jean-Paul Sartre*. Amherst: Massachusetts UP, 1991. x, 255 p. Cloth $29.95.

CODE, LORRAINE: *What Can She Know? Feminist Theory and the Construction of Knowledge*. Ithaca: Cornell, 1991. xiv, 349 p. Cloth $42.50, paper $14.95.

CRANSTON, MAURICE: *The Noble Savage: Jean-Jacques Rousseau 1754–1762*. Chicago: University Press, 1991. xiv, 399 p. Cloth $32.50.

CRITTENDEN, CHARLES: *Unreality: The Metaphysics of Fictional Objects*. Ithaca: Cornell, 1991. xiii, 177 p. Cloth $26.50, paper $9.95.

CURRIE, GREGORY: *The Nature of Fiction*. New York: Cambridge, 1990. xii, 222 p. Cloth $39.50.

DORE, CLEMENT: *Moral Scepticism*. New York: St. Martin's, 1991. x, 137 p. Cloth $45.00.

DUFF, R. A.: *Intention, Agency, and Criminal Liability: Philosophy of Action and the Criminal Law*. Cambridge: Blackwell, 1990. xiv, 219 p. Cloth $44.95, paper $15.95.

DUMMETT, MICHAEL: *Frege and Other Philosophers*. New York: Oxford, 1991. xiv, 330 p. Cloth $69.00.

EVANS, J. CLAUDE: *Strategies of Deconstruction: Derrida and the Myth of the Voice*. Minneapolis: Minnesota UP, 1991. xxiii, 205 p. Cloth $39.95, paper $14.95.

FONTANA, BIANCAMARIA: *Benjamin Constant and the Post-Revolutionary Mind*. New Haven: Yale, 1991. xvii, 165 p. Cloth $25.00.

FRIEDMAN, WILLIAM J.: *About Time: Inventing the Fourth Dimension*. Cambridge: MIT, 1990. x, 147 p. Cloth $19.95.

GALSTON, MIRIAM: *Politics and Excellence: The Political Philosophy of Alfarabi*. Princeton: University Press, 1990. xi, 240 p. Cloth $35.00.

GOODMAN, RUSSELL B.: *American Philosophy and the Romantic Tradition*. New York: Cambridge, 1991. xi, 162 p. Cloth $34.50.

GOULD, THOMAS: *The Ancient Quarrel between Poetry and Philosophy*. Princeton: University Press, 1990. xxvii, 318 p. Cloth $39.50.

GRIFFIN, NICHOLAS: *Russell's Idealist Apprenticeship*. New York: Oxford, 1991. xiii, 410 p. Cloth $98.00.

HANNAY, ALASTAIR: *Human Consciousness*. New York: Routledge, 1990. x, 221 p. Cloth $47.50.

HARDING, SANDRA: *Whose Science? Whose Knowledge? Thinking from Women's Lives*. Ithaca: Cornell, 1991. xii, 319 p. Cloth $34.50, paper $12.95.

HARRIS, ZELLIG: *A Theory of Language and Information: A Mathematical Approach*. New York: Oxford, 1991. xii, 428 p. Cloth $69.00.

HILL, CHRISTOPHER S.: *Sensations: A Defense of Type Materialism*. New York: Cambridge, 1991. ix, 253 p. Cloth $44.50, paper $15.95.

HUNLEY, J. D.: *The Life and Thought of Friedrich Engels: A Reinterpretation*. New Haven: Yale, 1991. xiii, 184 p. Cloth $22.50.

ISHIGURO, HIDE: *Leibnitz's Philosophy of Logic and Language*. 2nd ed. New York: Cambridge, 1991. x, 246 p. Cloth $49.50, paper $16.95.

JOHNSON, LAWRENCE E.: *A Morally Deep World: An Essay on Moral Significance and Environmental Ethics*. New York: Cambridge, 1991. ix, 301 p. Cloth $39.50.

KADISH, MORTIMER R.: *Toward an Ethic of Higher Education*. Stanford: University Press, 1991. ix, 205 p. Cloth $29.50.

KYMLICKA, WILL: *Liberalism, Community, and Culture*. New York: Oxford, 1991. 280 p. Paper $19.95.

LEFTOW, BRIAN: *Time and Eternity*. Ithaca: Cornell, 1991. xii, 377 p. Cloth $42.95.

LEONARD, STEPHEN T.: *Critical Theory in Political Practice*. Princeton: University Press, 1990. xxiv, 295 p. Cloth $35.00.

LUCAS, J. R., and P. E. HODGSON: *Spacetime and Electromagnetism: An Essay on the Philosophy of the Special Theory of Relativity*. New York: Oxford, 1990. xv, 310 p. Cloth $63.00, paper $29.95.

MCCARTHY, THOMAS: *Ideals and Illusions: On Reconstruction and Deconstruction in Contemporary Critical Theory*. Cambridge: MIT, 1991. x, 254 p. Cloth $22.50.

MILLAR, ALAN: *Reasons and Experience*. New York: Oxford, 1991. x, 227 p. Cloth $69.00.

NIELSEN, KAI: *After the Demise of the Tradition: Rorty, Critical Theory, and the Fate of Philosophy*. Boulder, CO: Westview, 1991. x, 278 p. Cloth $38.50.

NORTON, DAVID L.: *Democracy and Moral Development: A Politics of Virtue*. Berkeley: California UP, 1991. xvi, 198 p. Cloth $29.95.

PERNER, JOSEF: *Understanding the Representational Mind*. Cambridge: MIT, 1991. xiv, 348 p. Cloth $35.00.

PHILLIPS, ANNE: *Engendering Democracy*. University Park: Pennsylvania State UP, 1991. 183 p. Cloth $28.50, paper $13.95.

PLATTS, MARK: *Moral Realities: An Essay in Philosophical Psychology*. New York: Routledge, 1991. viii, 232 p. Cloth $74.50.

PYBUS, ELIZABETH: *Human Goodness: Generosity and Courage*. Toronto: University Press, 1991. x, 149 p. Cloth $40.00.

QUINE, W. V., and RUDOLF CARNAP: *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work*. RICHARD CREATH, ed. Berkeley: California UP, 1991. xv, 484 p. Cloth $34.95.

REGAN, TOM: *The Thee Generation: Reflections on the Coming Revolution*. Philadelphia: Temple, 1991. xi, 180 p. Cloth $24.95.

ROWE, WILLIAM L.: *Thomas Reid on Freedom and Morality*. Ithaca: Cornell, 1991. x, 189 p. Cloth $26.95.

SCHLESINGER, GEORGE: *The Sweep of Probability*. Notre Dame: University Press, 1991. viii, 228 p. Cloth $28.95.

SMITH, P. CHRISTOPHER: *Hermeneutics and Human Finitude: Toward a Theory of Ethical Understanding*. New York: Fordham, 1991. xxviii, 291 p. Cloth $40.00.

SNARE, FRANCIS: *Morals, Motivation, and Convention: Hume's Influential Doctrines*. New York: Cambridge, 1991. xii, 322 p. Cloth $44.50.

SORELL, TOM: *Scientism: Philosophy and the Infatuation with Science*. New York: Routledge, 1991. xi, 206 p. Cloth $45.00.

SOSA, ERNEST: *Knowledge in Perspective: Selected Essays in Epistemology*. New York: Cambridge, 1991. xi, 250 p. Cloth $49.50, paper $16.95.

SPRAGENS, THOMAS A.: *Reason and Democracy*. Durham: Duke, 1990. xi, 281 p. Cloth $39.50, paper $17.95.

TILGHMAN, B. R.: *Wittgenstein, Ethics and Aesthetics: The View from Eternity*. Albany: SUNY, 1991. xiv, 193 p. Cloth $49.50, paper $16.95.

WEATHERFORD, ROY: *The Implications of Determinism*. New York: Routledge, 1991. x, 275 p. Cloth $45.00.

WESTBROOK, ROBERT B.: *John Dewey and American Democracy*. Ithaca: Cornell, 1991. xix, 570 p. Cloth $29.95.

WETTSTEIN, HOWARD: *Has Semantics Rested on a Mistake? And Other Essays*. Stanford: University Press, 1991. xi, 232 p. Cloth $29.50.

YOLTON, JOHN W.: *Locke and French Materialism*. New York: Oxford, 1991. vii, 239 p. Cloth $55.00.

ZUIDERVAART, LAMBERT: *Adorno's Aesthetic Theory: The Redemption of Illusion*. Cambridge: MIT, 1991. xxiv, 380 p. Cloth $35.00.

[illegible]

Gaston Bachelard, Subversive Humanist
Texts and Readings
MARY McALLESTER JONES

McAllester Jones provides a generous sampling of Bachelard's seminal writings about science, poetry, and human consciousness. She frames her translations with critical essays on the development of Bachelard's ideas and his significant relationship to the contemporary French intellectual revolution commonly associated with Michel Foucault and Jacques Derrida.

Science and Literature series—George Levine, General Editor
Paper $15.00 Library cloth edition $37.50

The University of Wisconsin Press
114 North Murray Street • Madison, WI 53715 • (608) 262-2994
Wisconsin

Philosophy
from
ROUTLEDGE

In Praise of the Cognitive Emotions
And Other Essays in the Philosophy of Education

Israel Scheffler

In this intriguing collection of his most recent essays, Scheffler offers a critique of computer applications in education, develops a notion of strategic rationality in understanding mathematical reasoning, and contrary to prevalent notions of moral education, connects reason with care, emphasizing the intimate connection between emotion and reason and challenging the dominant perception of the two as oppositional.

192 pp $39.95/cloth $13.95/paper

Gadamer and Hermeneutics
Edited by **Hugh J. Silverman**

Gadamer and Hermeneutics provides a series of essays debating the virtues of Gadamerian hermeneutics versus aspects of the work of Heidegger, Ricoeur, Barthes, Derrida and Habermas, with each essay shedding special light on the hermeneutic process.

288 pp $45.00/cloth $15.95/paper

Truth and Eros
Foucault, Lacan and the Question of Ethics

John Rajchman

In ***Truth and Eros***, John Rajchman isolates the question of ethics in the work of Foucault and Lacan and explores its ramifications and implications for the present day.

192 pp $42.50/cloth $13.95/paper

Death and Desire
Psychoanalytic Theory in Lacan's Return to Freud

Richard Boothby

Death and Desire presents in Lacanian terms a new integration of psychoanalytic theory in which the battery of key Freudian concepts—from the dynamics of the Oedipus complex to the topography of ego, id and superego—are seen to intersect in Freud's most far-reaching and speculative formulation of a drive toward death.

256 pp $42.50/cloth $15.95/paper

Writing and the Moral Self
Berel Lang

"No one writes with more care and decency about a language than Berel Lang. ***Writing and the Moral Self*** maintains a high level of discrimination throughout, whether Lang is writing about gender pronouns or the Nazi genocide. It is a civilized and civilizing book." —Richard Ohmann, Wesleyan University

160 pp $42.50/cloth $13.95/paper

Michel Foucault, Philosopher
Edited and translated by
T. J. Armstrong
Preface by Lawrence D. Kritzman

Michel Foucault, Philosopher brings together original essays by an outstanding group of international scholars who present an array of critical readings of Foucault's work and its impact on western thought.

368 pp $49.95/cloth $16.95/paper

ROUTLEDGE
An imprint of Routledge, Chapman & Hall
29 West 35th Street
New York, New York 10001

Classics in Philosophy

GREEK PHILOSOPHY
2nd Edition
Reginald E. Allen
ISBN 0-02-900660-0 $12.95 paper

POLITICAL IDEAS OF ST. THOMAS AQUINAS
St. Thomas Aquinas
ISBN 0-02-840380-0 $10.95 paper

EIGHTEENTH CENTURY PHILOSOPHY
Lewis W. Beck, Editor
ISBN 0-02-902100-6 $12.95 paper

A MODERN INTRODUCTION TO PHILOSOPHY, 3rd Edition
Paul Edwards & Arthur Pap, Editors
ISBN 0-02-909200-0 $27.95

A SHORT HISTORY OF CHINESE PHILOSOPHY
Fung Yu-Lan
ISBN 0-02-910980-9 $14.95 paper

NINETEENTH-CENTURY PHILOSOPHY
Patrick Gardiner, Editor
ISBN 0-02-911220-6 $12.95 paper

INTRODUCTORY READINGS IN AESTHETICS
John Hospers
ISBN 0-02-915260-7 $14.95 paper

MORAL AND POLITICAL PHILOSOPHY
David Hume
ISBN 0-02-846170-3 $12.95 paper

DIALOGUES CONCERNING NATURAL RELIGION
David Hume
ISBN 0-02-846180-0 $10.95 paper

ESSAYS IN PRAGMATISM
William James
ISBN 0-02-847140-7 $10.95 paper

CRITIQUE OF JUDGMENT
Immanuel Kant
ISBN 0-02-847500-3 $12.95 paper

JOHN STUART MILL'S PHILOSOPHY OF SCIENTIFIC METHOD
John Stuart Mill
ISBN: 0-02-849250-1 $12.95 paper

PHILOSOPHY AND TECHNOLOGY
Carl Mitcham and
Robert Mackey, Editors
ISBN: 0-02-921430-0 $14.95 paper

THE SPIRIT OF THE LAWS
Baron de Montesquieu
ISBN 0-02-849270-6 $13.95 paper

NEWTON'S PHILOSOPHY OF NATURE
Isaac Newton
ISBN 0-02-849700-7 $12.95 paper

A CRITICAL HISTORY OF WESTERN PHILOSOPHY
D.J. O'Connor
ISBN 0-02-923840-4 $14.95 paper

PHILOSOPHY OF THE SIXTEENTH AND SEVENTEENTH CENTURIES
Richard H. Popkin
ISBN 0-02-925490-6 $12.95 paper

GREEK AND ROMAN PHILOSOPHY AFTER ARISTOTLE
Jason L. Saunders
ISBN 0-02-927730-2 $14.95 paper

ETHICS
Benedict de Spinoza
ISBN 0-02-852650-3 $10.95 paper

TWENTIETH CENTURY PHILOSOPHY
The Analytic Tradition
Morris Weitz
ISBN 0-02-934990-7 $12.95 paper

MEDIEVAL PHILOSOPHY
John F. Wippel & Allan B. Wolter
ISBN: 0-02-935650-4 $12.95 paper

Attention College Professors: Examination copies will be sent on a 30-day approval basis only. Simply place your request on university stationery, indicating book requested, course title, projected enrollment, and date your next decision is due. When ordering examination copies of paperbacks, enclose a check for $2.00 per copy to cover postage and handling. Mail to: The Free Press, Attn: Lisa Weinberg, 866 Third Avenue, NY, NY 10022

Fp **THE FREE PRESS**
A Division of Macmillan, Inc.
866 Third Avenue, New York, New York 10022

PHILOSOPHY PAST AND PRESENT

NEW from PENGUIN USA

LUDWIG WITTGENSTEIN
The Duty of Genius
Ray Monk

A unique insight into the life and work of Wittgenstein, one of the most acute philosophical minds of the twentieth century. "Well written, intelligent and moving...a first-rate biography that stands as both a fine piece of scholarship and a vivid portrait."—*The New York Times.* 672 pp.

Penguin paperback
0-14-015995-9 $15.95
Available November 1991

A NEW VIEW OF SOCIETY AND OTHER WRITINGS
Robert Owen

Edited and Introduced by Gregory Claeys. A diverse selection of inspiring writings from Robert Owen, one of the greatest British social reformers, reflecting his intense concern for equality and justice through his radical views of British society, working practices, education, and the Poor Laws. 416 pp.

Penguin Classic paperback
0-14-043348-1 $9.95

PERSONALITY
An Introduction to The Theories of Psychology
Peter Morea

In this accessible volume, Morea examines and explains the main theories of personality through a survey of the well-established approaches of Freud, Jung, Skinner, and others, providing a comprehensive survey of their insights and ideas and an objective and multi-sided understanding of the subject. 192 pp.

Penguin paperback
0-14-012519-1 $7.95

THE ENNEADS
Plotinus

Translated by Stephen MacKenna. Edited by John Dillon. This interpretative translation of Plotinus' highly original synthesis of Platonism, mystic passion, Greek philosophy, and variants of the Trinity and other central doctrines offers profound insights into an often inaccessible yet brilliant thinker. 688 pp.

Penguin Classic paperback
0-14-044520-X $12.95
Available November 1991

THE FIRE AND THE SUN
Why Plato Banished the Artists
Iris Murdoch

Based on her 1976 Romanes Lecture at Oxford University, Murdoch's latest is a succinct, witty book which discusses Plato's attitude toward art and the age-old quarrel between poetry and philosophy. Providing an accessible and necessary compression of Plato's thought, *The Fire and the Sun* also includes observations on Kant, Tolstoy, Freud, and Kierkegaard. 96 pp.

Viking hardcover
0-670-83708-3 $14.95

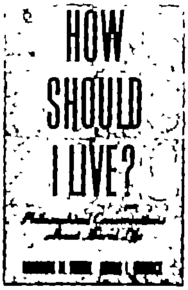
JOURNEY THROUGH GENIUS
The Great Theorems of Mathematics
William Dunham

This rare combination of the historical, biographical, mathematical, and philosophical offers a captivating introduction to a neglected, important field of human creativity. "A fascinating lingering over individual examples of ingenuity and insight."—Isaac Asimov. 448 pp.

Penguin paperback
0-14-014739-X $10.95

PENGUIN USA

Academic Marketing Department, 375 Hudson Street, New York, NY 10014-3657



Outstanding New Texts

HOW SHOULD I LIVE?

Philosophical Conversations About Moral Life

RANDOLPH M. FEEZELL & CURTIS L. HANCOCK

Eight lively, fascinating, fictional dialogues on everything from the definition to the value of ethics. Critical discussions of many prevailing theories including divine command theory, relativism, egoism, utilitarianism, and deontology. The book also contains treatments of a natural law approach to virtue-ethics and Carol Gilligan's widely discussed account of female moral development in the context of an ethic of care. Every dialogue ends with a list of key terms, epigrams, and suggested readings. "An extremely practical book which should prove helpful to many philosophy professors." —Ralph McInerny, University of Notre Dame

218 pp., Paper, ISBN 1-55778-284-9, $14.95

PHILOSOPHY

An Introduction Through Literature

LOWELL KLEIMAN & STEPHEN LEWIS

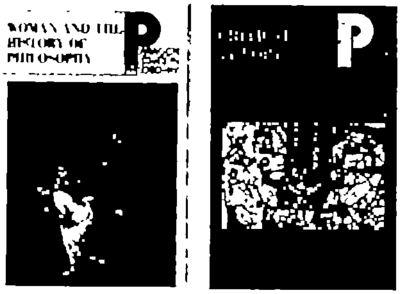
Selections from the world's greatest philosophers and authors in a thoughtfully designed anthology that effectively uses literary works to illustrate abstract points. The text is organized around the seven basic principles that are generally covered in introductory courses: knowledge; truth; personal identity; ethics; justice; religious belief; freedom, fatalism, and determinism.

Selections have been chosen so the relationship between the literary piece and the philosophical issue under discussion flows easily. Complete short stories rather than excerpts are used so students gain the nuances and subtleties evident only in an entire work.

"A most welcome and enlightening addition."—*AITIA Philosophy-Humanities Magazine*

618 pp., Cloth, ISBN 1-55778-171-0, $32.00

ATTENTION PROFESSORS: To obtain examination copies, send your request on school letterhead, indicating your name and academic position, course name and number, projected enrollment, and date your adoption decision is due.



from Paragon House

PARAGON ISSUES IN PHILOSOPHY

This new and innovative series meets the needs of today's students and professors with undergraduate texts and readers by a team of distinguished philosophers. As well as contributing to the contemporary discussion on traditional issues in philosophy, the series breaks new ground in up-to-the-minute specialized areas such as ethnicity, sport, critical theory, and feminist scholarship.

WOMAN AND THE HISTORY OF PHILOSOPHY

NANCY TUANA

Using the works of Aristotle, Plato, Descartes, Rousseau, Kant, Hume, Locke, and Hegel, the author offers a detailed examination of the history of philosophical theories concerning woman's nature and traits viewed as feminine. Topics covered include: the second sex; women and reason; women and morality; private woman, public man; the future of "feminine" and "masculine." An excellent and important text for undergraduate students.

140 pp., Paper, ISBN 1-55778-194-X, $14.95

CRITICAL THEORY

The Essential Readings

DAVID INGRAM & JULIA SIMON-INGRAM

Selections from the most important works of the Critical Theorists—Theodor Adorno, Max Horkheimer, Herbert Marcuse, and Jurgen Habermas—followed by a critique of their work by prominent post-structuralists, post-modernists, and feminists. A companion volume to *Critical Theory and Philosophy* and suitable for advanced scholars as well as beginning students.

388 pp., Paper, ISBN 1-55778-355-5, $15.95

PARAGON HOUSE

90 FIFTH AVENUE, NEW YORK, NY 10011/*1-800-PARAGON*

Cambridge University Press

The Political Responsibility of Intellectuals

Ian Maclean, Alan Montefiore, and Peter Winch, Editors

The Political Responsibility of Intellectuals addresses the many problems in defining the relationship of intellectuals to the society in which they live. The contributors seek to provide tentative answers to the questions posed.

Contributors: *Alan Montefiore, Peter Winch, Ernest Gellner, Elemer Hankiss, Jerzy Jedlicki, Jacek Kurczewski, Jean-Jacques Lecercle, David J Levy, Graham Lock, Ian Maclean,* [illegible] *Jerzy Szacki, G M. Tamas, Edward Shils.*

39179-2 Hardcover $49.50
39859-2 Paper $15.95

Morals, Motivation, and Convention

Hume's Influential Doctrines

Francis Snare

"…there is substantial contemporary interest in the Humean doctrines and arguments which Snare scrupulously lays out and criticizes. Many people have discussed these themes in various ways, but I don't know of a clearer, more careful, and sounder treatment of them than Snare's."

—*Stephen Darwall, University of Michigan*

Cambridge Studies in Philosophy

39261-6 Hardcover $44 50

Knowledge in Perspective

Selected Essays in Epistemology

Ernest Sosa

"Sosa's epistemology synthesizes rival perspectives in an extremely subtle and sophisticated fashion. This is a very welcome collection." —*Alvin I. Goldman, The University of Arizona*

35628-8 Hardcover $49 50
39643-3 Paper $16 95

Probabilistic Causality

Ellery Eells

In this important new book Ellery Eells explores and refines current philosophical conceptions of probabilistic causality. A salient feature of the book is a new theory of token level probabilistic causation in which the evolution of the probability of a later event from an earlier event is central.

Cambridge Studies in Probability, Induction and Decision Theory

39244-6 Hardcover $44.50

Liability and Responsibility

Essays in Law and Morals

R. G. Frey and Christopher W. Morris, Editors

This collection of contemporary essays by a group of well-known philosophers and legal theorists covers various topics in the philosophy of law; focusing on issues concerning liability in contract, tort, and criminal law.

Contributors: *Randy E Barnett, Jules L Coleman, Richard A Epstein, John Finnis, George P Fletcher, Robert E Goodin, Jean Hampton, Douglas D Heckathorn, Anthony Kenny, Steven M. Massey, Jeffrie G Murphy, Glen O. Robinson, Ernest J Weinrib, Alan R. White*

Cambridge Studies in Philosophy and Law

39216-0 Hardcover $54 50

New in paperback.

Rules, Norms, and Decisions

On the Conditions of Practical and Legal Reasoning in International Relations and Domestic Affairs

Friedrich V. Kratochwil

By assessing the impact of norms on decision making, this book argues that norms influence choices by providing reasons rather than by being causes for action. It approaches the problem via an investigation of the reasoning process in which norms play a decisive role.

Cambridge Studies in International Relations 2

40971-3 Paper $16 95

Moral Legislation

A Legal-Political Model for Indirect Consequentialist Reasoning

Conrad D. Johnson

This is a book about moral reasoning: how we actually reason and how we ought to reason. It defends a form of "rule" utilitarianism whereby we must sometimes judge and act in moral questions in accordance with generally accepted rules, so long as the existence of those rules is justified by the good they bring about.

Cambridge Studies in Philosophy

39224-1 Hardcover $39.50

Available in bookstores or write:

CAMBRIDGE
UNIVERSITY PRESS
40 W. 20th St., N.Y., NY 10011-4211.
Call toll-free 800-872-7423.
MasterCard/VISA accepted.
Prices subject to change.

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NUMBER 10
OCTOBER 1991

Special Issue, Eighty-eighth Annual Meeting
American Philosophical Association, Eastern Division

page

Symposium: Logical Positivism: A Retrospective

505 *The Re-evaluation of Logical Positivism* Michael Friedman
520 *Logical Positivism: A Retrospective* David Weissman
522 *Convention, Neutrality, and the Limits of Logic* Richard Creath

Symposium: Post-Kantian Idealism

524 *Why Fichte Now?* Daniel Breazeale
532 *Idealism and Agency in Kant and Hegel* Robert B. Pippin
542 *Fichte: Foundationalism, Antifoundationalism, and the New Nihilism* Chernor M. Jalloh

Symposium: Is Music an Art?

544 *Is Music an Art?* Peter Kivy
555 *The Arts of Music* Philip Alperson

Symposium: Philosophy in a Different Voice

557 *Philosophy in a Different Voice* Janet A. Kourany
568 *Feminism and the Canon* Pamela Hall

570 Johnsonian Prize in Philosophy
571 Program of the Meetings
585 List of Program Participants
590 Appendix: Programs of Group Meetings
601 List of Group Participants

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1991 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER. Second class postage paid at New York, NY, and additional mailing offices. Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NO. 10, OCTOBER 1991

THE RE-EVALUATION OF LOGICAL POSITIVISM*

It is now well over half a century since the heyday of the philosophical movement known as logical positivism or logical empiricism. Depending on how one counts, it is now approaching half a century since the official demise of this movement.[1] Since that demise it has naturally been customary to view logical positivism as a kind of philosophical bogeyman whose faults and failings need to be enumerated (or, less commonly, investigated) before one's favored "new" approach to philosophy can properly begin. And such an attitude toward logical positivism and its demise has been widely prevalent, not only in the narrower community of philosophers of science (who have characteristically proceeded against the background of Thomas Kuhn's well-known critique), but also in the broader philosophical community as well. With our increasing historical distance from logical positivism, however, a more dispassionate attitude has also inevitably begun to emerge. No longer threatened or challenged by logical positivism as a live philosophical option, it is becoming increasingly possible to consider this movement as simply a part of the history of philosophy which, as such, can be investigated impartially from a historical point of view. Indeed, we have seen in recent years a veritable flowering of historically oriented reconsiderations of logical positivism.[2]

---

* To be presented in an APA symposium on Logical Positivism, December 28. David Weissman and Richard Creath will comment. See this JOURNAL, this issue, 520–1 and 522–3, respectively, for their contributions.

[1] I assume that this event took place sometime between the publication of W. V. Quine's "Two Dogmas of Empiricism" (1951), and that of Thomas Kuhn's *The Structure of Scientific Revolutions* (1962).

[2] I have in mind, for example, work in this country by A. Coffa, R. Creath, B. Dreben, W. Goldfarb, D. Howard, J. Lewis, T. Oberdan, A. Richardson, T Ricketts, T. Ryckman, H. Stein, T. Uebel, Z. Zhai; work in Great Britain by S. Haack, A. Hamilton, J. Skorupski; work on the Continent by H.-J. Dahms, R. Haller, R.

0022-362X/91/8810/505–19

In the course of these reconsiderations it has become clear—not at all surprisingly, of course—that the above-mentioned post-positivist reaction gave birth to a large number of seriously misleading ideas about the origins, motivations, and true philosophical aims of the positivist movement. (One can hardly expect philosophical critics, concerned largely with their own agendas rather than with historical fidelity, to generate anything other than stereotypes and misconceptions.) I shall discuss what I take to be some of the most important of such misleading ideas in what follows. But I also hope to show—or at least to suggest—that achieving a better understanding of the background, development, and actual philosophical context of logical positivism is not merely of historical interest. For the fact remains that our present situation evolves directly—for better or for worse—from the rise and fall of positivism, and what I want to suggest is that we will never successfully move beyond our present philosophical situation until we attain a properly self-conscious appreciation of our own immediate historical background.

I

Perhaps the most misleading of the stereotypical characterizations views logical positivism as a version of philosophical "foundationalism." The positivists—so this story goes—were concerned above all to provide a philosophical justification of scientific knowledge from some privileged, Archimedean vantage point situated somehow outside of, above, or beyond the actual (historical) sciences themselves. More specifically, they followed the lead of the logicist reduction of mathematics to logic—where the latter is also understood as fundamentally foundationalist in motivation and import. Just as the logicists attempted to justify mathematical knowledge and place it on a secure foundation by means of a derivation from (supposedly more certain) logical knowledge, so the positivists attempted to justify empirical science and place it on a secure foundation by logically constructing the concepts of empirical science on the basis of the (supposedly more certain) immediate data of sense. Thus, formal logic furnished the foundational enterprise with the required Archimedean standpoint located outside of the actual (historical) sciences themselves, and phenomenalist reductionism carried out rigorously

---

Hegselmann, M. Heidelberger, K. Hentschel, A. Kamlah, D. Koppelberg, U. Majer, V. Mayer, C Moulines, P Parrini, J. Proust, H. Rutte, W. Sauer, W. Vossenkuhl, G. Wolters. In addition, the *Vienna Circle Collection*, under the general editorship of R. Cohen, has recently made available in English many previously untranslated works of Schlick, Neurath, Reichenbach, Menger, and Waismann. Carnap's earlier works are unfortunately not yet available in English translation.

using the tools of formal logic (as epitomized in Carnap's *Der logische Aufbau der Welt*[3] of 1928) then provided the desired epistemological justification of the sciences.

This conception of the aims and posture of philosophy vis-à-vis the special sciences represents an almost total perversion of the actual attitude of the logical positivists, who rather considered their intellectual starting point to be a *rejection* of all such philosophical pretensions. An eloquent example is found in the first paragraph of a paper by Schlick[4] from 1915 on the need for philosophy to adapt itself to the new findings of relativity theory:

> We have known since the days of Kant that the only fruitful method of all theoretical philosophy consists in critical inquiry into the ultimate principles of the special sciences. Every change in these ultimate axioms, every emergence of a new fundamental principle, must therefore set philosophical activity in motion, and has naturally done so even before Kant. The most brilliant example is doubtless the birth of modern philosophy from the scientific discoveries of the Renaissance. And the Kantian Critical Philosophy may itself be regarded as a product of the Newtonian doctrine of nature. It is primarily, or even exclusively, the principles of the exact sciences that are of major philosophical importance, for the simple reason that in these disciplines alone do we find foundations so firm and sharply defined, that a change in them produces a notable upheaval, which can then also acquire an influence on our world-view (*ibid.*, p. 153).

Schlick goes on to argue that neither of the two prevailing philosophical systems—neither the "neo-Kantianism" of Cassirer and Natorp nor the "positivism" of Petzoldt and Mach—can do justice to Einstein's new theory and *therefore both systems must be aban-*

---

[3] *The Logical Structure of the World*, R. George, trans. (Berkeley: California UP, 1967) [hereafter *Aufbau*].

The most explicit example of this kind of characterization of logical positivism of which I am aware is found in R. Giere, *Explaining Science: A Cognitive Approach* (Chicago: University Press, 1988), pp. 22–8. This kind of conception is certainly aided and abetted (although with important reservations in each case) by Quine's "Epistemology Naturalized" (1968) and A. J. Ayer's *Language, Truth and Logic* (1936). Richard Rorty has recently portrayed the entire modern philosophical tradition as based on "foundationalist epistemology," under which rubric the logical positivists are also clearly thought to fall: see *Philosophy and the Mirror of Nature* (Princeton: University Press, 1979), pp. 59, 332–3 (and it is thus clear for Rorty that Kuhn's critique is predicated on a rejection of such foundationalist epistemology).

[4] "Die philosophische Bedeutung des Relativitätsprinzips," *Zeitschrift für Philosophie und philosophische Kritik*, CLIX (1915): 129–75—P. Heath, trans. "The Philosophical Significance of the Principle of Relativity," in H. Mulder and B. van de Velde-Schlick, eds., *Moritz Schlick: Philosophical Papers* (Dordrecht: Reidel, 1979), Vol. I (part of the above mentioned *Vienna Circle Collection*).

doned. Entirely new philosophical principles—based on the work of Einstein himself and of Poincaré—are necessarily required.

For Schlick, then, philosophy as a discipline is in no way foundational with respect to the special sciences. On the contrary, it is the special sciences that are foundational for philosophy. The special sciences—more specifically, the "exact sciences"—are simply taken for granted as paradigmatic of knowledge and certainty. Far from being in a position somehow to justify these sciences from some higher vantage point, it is rather philosophy itself that is inevitably in question. Philosophy, that is, must *follow* the evolution of the special sciences so as to test itself and, if need be, to reorient itself with respect to the far more certain and secure results of these sciences. In particular, then, the central problem of philosophy is not to provide an epistemological foundation for the special sciences (they already have all the foundation they need), but rather to redefine its own task in the light of the recent revolutionary scientific advances that have made all previous philosophies untenable.[5]

Moreover, this conception of the proper stance of philosophy vis-à-vis the special sciences is not peculiar to Schlick, it is in fact characteristic of the logical positivists generally. This is so even—and indeed especially—of Carnap's *Aufbau* (which work is, of course, standardly supposed to be most representative of foundationalist epistemology). It is true that the *Aufbau* presents a phenomenalist reduction of all concepts of science to the immediately given data of experience. Yet the point of this construction has little if anything to do with traditional foundationalism. First of all, Carnap shows no interest whatever in the philosophical skepticism that motivates (for example) Russell in *Our Knowledge of the External World* (1914), nor does the text of the *Aufbau* at any point engage the traditional vocabulary of "certainty," "doubt," "justification," and so on.[6] Secondly, and more importantly, Carnap is perfectly explicit that the particular constructions he employs depend entirely on the actual "results of the empirical sciences" (§122) and, accordingly, that the particular constructional system he presents is best viewed as a "rational reconstruction" of the actual (empirical) process of cognition (§100).[7] Thus, for example, Carnap's choice of holistic "elementary experiences" rather than atomistic sensations as the basis of his system is grounded in the empirical findings of Gestalt psychology (§67), his definition of the visual sense modality depends on the (supposed) empirical fact that it is the unique sense modality having exactly five dimensions (§86), and so on. In particular, then, the findings of the special sciences (such as empirical psychology) are in no way in question in the *Aufbau:* once again, it is philosophy that must adapt itself to them rather than the other way around.

---

[5] As Schlick indicates, this is also a fruitful way of viewing Kant's philosophical stance with respect to the special sciences. For Kant, too, it is philosophy itself rather than the special sciences which is in question and which therefore requires justification. Kant's aim is not to ground the sciences in something firmer and more secure (for it is they that are paradigmatic of certainty) but rather to reform *metaphysics* in accordance with the already achieved successes of the exact sciences. See the "Preface" to the second edition of the *Critique of Pure Reason* (especially Bxxii–xxiii), §VI of the "Introduction," and §40 of the *Prolegomena* This undercuts decisively, it seems to me, Rorty's portrayal (*op. cit.*) of Kant as the *echt*-foundationalist.

[6] Carnap does employ this language in retrospectively describing the motivation of the *Aufbau* at several points in his "Intellectual Autobiography," in P. Schilpp, ed., *The Philosophy of Rudolf Carnap* (La Salle: Open Court, 1963), pp. 51, 57. Nevertheless, the contrast between this retrospective account and the language of the *Aufbau* itself is striking indeed.

The aim of the *Aufbau,* therefore, is not to use logic together with sense data to provide empirical knowledge with an otherwise missing epistemological foundation or justification. Its aim, rather, is to use recent advances in the science of logic (in this case, the Russellian type theory of *Principia Mathematica*) together with advances in the empirical sciences (Gestalt psychology, in particular) to fashion a scientifically respectable *replacement* for traditional epistemology. Carnap's depiction of the construction of scientific knowledge from elementary experiences via the logical techniques of *Principia Mathematica* enables us to avoid the metaphysical excesses of the traditional epistemological schools—"realism," "idealism," "phenomenalism," "transcendental idealism" (§177)—while, at the same time, capturing what is correct in all of these schools: allowing us to represent, in Carnap's words, the "neutral basis" (*neutrale Fundament*) common to all (§178).[8]

II

According to the standard picture of logical positivism briefly sketched above, that movement is to be understood not only as a species of foundationalist epistemology but also as a version of *empiricist* epistemology in the tradition of Locke, Berkeley, Hume, Mach, and Russell's external-world program.[9] The immediately given data of sense are viewed as the primary examplars of knowledge and certainty, and all other putative claims to knowledge are judged to be warranted—or, as the case may be, unwarranted—in light of their relations to the immediate data of sense. (When empiricist reductionism is acknowledged to have failed, the positivists' naive empiricism is then thought to express itself in the doctrine of an epistemically privileged and theory-neutral observation language against which all scientific claims are to be tested.) I have dealt with the conception of the positivists as foundationalists above, and this question is, I think, rather easily disposed of. The question of empiricism—and, in particular, of the relationship between the positivists and the traditional empiricism of Locke, Berkeley, Hume, and Mach—is, however, considerably more delicate.

---

[7] Cf. Schilpp, *op. cit.*, p. 18.

[8] For more on this kind of approach to the *Aufbau,* see my "Carnap's *Aufbau* Reconsidered," Noûs, xxi (1987): 521–45; A. Richardson, "How Not to Russell Carnap's *Aufbau,*" *PSA*, i (1990): 3–14, and A. Richardson, *Epistemology Purified: Objectivity, Logic, and Experience in Carnap's* Aufbau (Doctoral Dissertation: University of Illinois at Chicago, 1990).

[9] This is, of course, the conception promulgated especially in Ayer's *Language, Truth and Logic:* compare the first sentence of the "Preface" to the first edition, where the views put forward are characterized as "the logical outcome of the empiricism of Berkeley and David Hume."

The first point to notice is that the positivists' main philosophical concerns did not arise within the context of the empiricist philosophical tradition at all. The initial impetus for their philosophizing came rather from late nineteenth-century work on the foundations of geometry by Riemann, Helmholtz, Lie, Klein, and Hilbert—work which, for the early positivists, achieved its culmination in Einstein's theory of relativity.[10] The principal philosophical moral the early positivists drew from these geometrical developments was that the Kantian conception of pure intuition and the synthetic a priori could no longer be consistently maintained—consistently, that is, with the situation now presented to us by the exact sciences. In particular, Hilbert's logically rigorous axiomatization of Euclidean geometry (1899) shows conclusively that spatial intuition has no role to play in the reasoning and inferences of pure geometry, and the development of non-Euclidean geometries, together with their actual application to nature by Einstein, shows conclusively that our knowledge of geometry cannot be synthetic a priori in Kant's sense. All the early positivists were thus in agreement that the strictly Kantian conception of the a priori must be rejected, and this rejection of the synthetic a priori constituted a centrally important element in what they came to call their "empiricism."

---

[10] Most of the early writings of the positivists focused on these revolutionary mathematical-physical developments. In addition to the 1915 paper of Schlick cited above, see his *Raum und Zeit in der gegenwärtigen Physik* (1917)—*Space and Time in Contemporary Physics*, H. L. Brose, trans. (1920) reprinted in H. Mulder, *op. cit.*; H. Reichenbach, *Relativitätstheorie und Erkenntnis apriori* (1920)—*The Theory of Relativity and A Priori Knowledge*, M. Reichenbach, trans. (Berkeley: California UP, 1965); R. Carnap, *Der Raum: Ein Beitrag zur Wissenschaftslehre* Ergänzungsheft Nr. 56 of *Kantstudien* (1922)—*Space: A Contribution to the Theory of Science*, M. Friedman and P. Heath, trans., forthcoming. Interest in these themes was also stimulated by Weyl's 1919 edition of Riemann's 1854 dissertation and the 1921 edition of Helmholtz's geometrical writings by Schlick and the physicist P. Hertz—the latter is translated by M. Lowe as *Hermann von Helmholtz: Epistemological Writings*, R. Cohen and Y. Elkana, eds. (Dordrecht: Reidel, 1977), as part of the *Vienna Circle Collection*.

Yet it is equally important to notice, in the second place, that the positivists did not react to the demise of the Kantian synthetic a priori by adopting a straightforwardly empiricist conception of physical geometry of the kind traditionally imputed to Gauss (who is reported to have attempted to determine the curvature of physical space by measuring the angle sum of a terrestrial triangle determined by three mountain tops). On the contrary, all the early positivists also strongly rejected this kind of empiricist conception (which they attributed to Gauss, Riemann, and, at times, Helmholtz) and rather followed the example of Poincaré in maintaining that there is no direct route from sense experience to physical geometry: essentially nonempirical factors, variously termed "conventions" or "coordinating definitions," must necessarily intervene between sensible experience and geometrical theory. The upshot is that it is in no way a straightforward empirical matter of fact whether space is Euclidean or non-Euclidean.[11]

This radically new conception of physical geometry—neither strictly Kantian nor strictly empiricist—was formulated by Reichenbach in an especially striking fashion in his first book, *The Theory of Relativity and A Priori Knowledge* (1920). Reichenbach maintains a sharp distinction, within the context of any given scientific theory, between two intrinsically different types of principles: "axioms of connection" are empirical laws in the traditional sense recording inductive regularities involving terms and concepts that are already sufficiently well-defined; "axioms of coordination," on the other hand, are nonempirical statements that must be antecedently laid down before the relevant terms and concepts have a well-defined subject matter in the first place. (Thus, for example, Gauss's attempt empirically to determine the curvature of space via a terrestrial triangle inevitably fails, because it tacitly presupposes that light rays travel in straight lines and, therefore, that the notion of "straight line" is already well-defined—but how, independently of the geometrical and optical principles supposedly being tested, can this possibly be done?) These nonempirical axioms of coordination—which include, paradigmatically, the principles of physical geometry—are thus "constitutive of the object of knowledge," and, in this way, we can therefore vindicate *part* of Kant's conception of the a priori.

---

[11] See esp. *La Science et L'Hypothèse* (1902)—*Science and Hypothesis*, J. Larmor, trans. (New York. Dover, 1952); compare the famous remark (at the conclusion of ch. V.5): "Whichever way we look at it, it is impossible to discover in geometrical empiricism a rational meaning." Einstein himself frequently expressed admiration for Poincaré's doctrine: see, e.g , "Geometrie und Erfahrung" (1921)—"Geometry and Experience," G. Jeffrey and W. Perret, trans. in *Sidelights on Relativity* (New York: Dutton, 1923)

For, according to Reichenbach, the notion of aprioricity had two independent aspects in Kant: the first involves necessary and unrevisable validity, but the second involves only the just-mentioned feature of "constitutivity." The lesson of modern geometry and relativity theory, then, is not that the Kantian a priori must be abandoned completely, but rather that the constitutive aspect must be separated from the aspect of necessary validity. Physical geometry is indeed nonempirical and constitutive—it is not itself subject to straightforward observational confirmation and disconfirmation but rather first makes possible the confirmation and disconfirmation of properly empirical laws (viz., the axioms of connection). Nevertheless, physical geometry can still evolve and change in the transition from one theoretical framework to another: Euclidean geometry, for example, is a priori in this constitutive sense in the context of Newtonian physics, but only topology (sufficient to *admit* a Riemannian structure) is a priori in the context of general relativity.

Reichenbach concludes that traditional empiricism is in error in not recognizing the a priori constitutive role of axioms of coordination, and it is clear, moreover, that, despite some terminological wrangling on this point, the other logical positivists are in substantial agreement.[12] One of the central themes of Schlick's *General Theory of Knowledge* (1918), for example, is a sharp dichotomy between raw sensible acquaintance and genuine objective knowledge. Immediate contact with the given is both fleeting and irredeemably subjective. Objective knowledge therefore requires *concepts* and *judgments*, which are to be carefully distinguished from intuitive sensory presentations. Concepts and judgments are in fact only possible in the context of a rigorous formal system, of which Hilbert's axiomatization of geometry is paradigmatic. More specifically, the Hilbertian notion of "implicit definition" of scientific concepts via their logical places in a formal system (which notion is explicitly associated by Schlick with Poincaré's conventionalist philosophy of geometry) can alone explain how rigorous, exact, and truly objective representation is possible. We are here obviously very far from traditional empiricism and very close indeed to the supposedly antipositivist doctrine of the theory ladenness of observation.[13]

---

[12] The terminological wrangling in question arises in correspondence between Schlick and Reichenbach in November of 1920 concerning Reichenbach's book. Schlick criticizes Reichenbach for conceding too much to the Kantian side and argues that Reichenbach's a priori constitutive principles should rather be understood as conventions à la Poincaré. This correspondence is discussed in A. Coffa, *The Semantic Tradition from Kant to Carnap* (New York: Cambridge, 1991), ch. 10. Schlick himself alludes to this correspondence in "Kritizistische oder empiristische Deutung der neuen Physik?" *Kantstudien*, xxvi (1921)—"Critical or Empiricist Interpretation of Modern Physics?" P. Heath, trans., in H. Mulder, *op. cit.*

Once again, the same is true even—and indeed especially—of Carnap's *Aufbau*. Carnap outlines an elaborate construction of all scientific concepts from the immediately given data of sense, using the logical machinery of *Principia Mathematica*. Yet he also holds that the objective meaning of scientific concepts can in no way depend on merely ostensive contact with the given. On the contrary, intersubjective communication is possible only in virtue of the *logical structure* of the concepts in question arising from their logical places within the total system of scientific knowledge. More specifically, intersubjective meaning must derive entirely from what Carnap calls "purely structural definite descriptions" rather than from sensory ostension (§§12-5). (Thus, for example, the visual sense modality is defined in terms of the purely formal properties of its dimensionality rather than in terms of its phenomenal content.) In this sense, the elaborate logical structure erected above the basic elements of the system (the elementary experiences) is actually more important than the basic elements themselves. Objective meaning flows from the top down, as it were, rather than from the bottom up.[14]

Carnap illuminatingly articulates the precise meaning of the resulting "empiricism" in the "Preface" to the second edition (1961!) of the *Aufbau:*

> For a long time, philosophers of various persuasions have held the view that all concepts and judgments result from the cooperation of experience and reason. Basically, empiricists and rationalists agree in

---

[13] For further discussion of Schlick, see my "Moritz Schlick's *Philosophical Papers*," *Philosophy of Science*, L (1983): 498–514. Theory ladenness follows from the circumstance that *all* scientific concepts—including those used to report the results of experiments—acquire objective meaning from their places in a theoretical system. For further discussion of this point, see my "Philosophy and the Exact Sciences: Logical Positivism as a Case Study," in J. Earman, ed., *Empiricism, Explanation and the Philosophy of Science* (Berkeley: California UP, forthcoming).

[14] For further discussion see again the references cited in fn. 8 above.

> this view, even though both sides give a different estimation of the two factors, and obscure the essential agreement by carrying their viewpoints to extremes. The thesis which they have in common is frequently stated in the following simplified version: The senses provide the material of cognition, reason works up [*verarbeitet*] the material so as to produce an organized system of knowledge. There arises then the problem of finding a synthesis of traditional empiricism and traditional rationalism. Traditional empiricism rightly emphasized the contribution of the senses, but did not realize the importance of logical and mathematical forms. . . . I had realized, on the one hand, the fundamental importance of mathematics for the formation of a system of knowledge and, on the other hand, its purely logical, purely formal character to which it owes its independence from the contingencies of the real world. These insights formed the basis of my book. . . . This orientation is sometimes called "logical empiricism" (or "logical positivism"), in order to indicate the two components (*op. cit.*, pp. v–vi).

In thus emphasizing the central importance of a priori formal elements in first providing objective meaning for the otherwise undigested immediate data of sense, logical positivism has, it seems to me, broken decisively with the traditional empiricism of Locke, Berkeley, Hume, and Mach—which tradition is understood by the positivists themselves as mistakenly giving epistemic centrality to precisely such undigested immediate sensory data. Perhaps the best way to put the point is that the logical positivists have staked out an entirely novel position that is, as it were, intermediate between traditional Kantianism and traditional empiricism: it gives explicit recognition to the constitutive role of a priori principles, yet, at the same time, it also rejects the Kantian characterization of these principles as *synthetic* a priori.

III

The novel philosophical position briefly sketched above is, however, faced with several fundamental problems. The heart of the positivists' new conception is the idea of constitutive but nonsynthetic a priori principles underlying the possibility of genuinely objective scientific knowledge. There is a sharp distinction, in particular, between conventions, coordinating definitions, or axioms of coordination, on the one hand, and empirical principles properly so-called, on the other. On the one side lie the principles of pure mathematics and (at least in the context of some physical theories) the principles of physical geometry, on the other side lie standard empirical laws such as Maxwell's equations and the law of gravitation. But what is the basis for this distinction, and how, more generally, are we sharply to differentiate the two classes of principles? On a strictly Kantian view, such a distinction is, of course, grounded in the fixed constitution of our cognitive faculties, and the question of differentiation is correspondingly straightforward: the a priori principles are precisely those possessing necessary and unrevisable validity.[15] Now, however, we have explicitly acknowledged that a priori constitutive principles possess no such necessary validity and, in fact, that these principles may evolve and change—in response to empirical findings—with the progress of empirical science. So what exactly distinguishes our a priori principles from ordinary empirical laws properly so-called?

A second problem is perhaps even more fundamental. The logical positivists, I have argued, strongly rejected a foundationalist conception of philosophy vis-à-vis the special sciences. There is no privileged vantage point from which philosophy can pass epistemic judgment on the special sciences: philosophy is conceived as rather following the special sciences so as to reorient itself in response to their established results. But what then *is* the peculiar task of philosophy, and how, in general, does it relate to the special sciences? Is philosophy itself simply one special science among others, and, if not, from what perspective does it then respond to and rationally reconstruct the results of the special sciences? The positivists are nearly unanimous in explicitly rejecting a naturalistic conception of philosophy as simply one empirical science among others—a branch of psychology, perhaps, or of the sociology of knowledge.[16] On the whole, they instead prefer to think of philosophy as in some sense a branch of *logic* and to conceive the peculiarly philosophical task as that of "logical analysis" of the special sciences. Yet the perspective or point of view from which such logical analysis is to proceed remains radically unclear.

---

[15] Cf. Kant's well-known "criterion by which to distinguish with certainty between pure and empirical knowledge" articulated in §II of the "Introduction" to the *Critique of Pure Reason.*

[16] The conspicuous exception here is Neurath, who articulates a version of naturalism bearing some similarities to well-known views of Quine's. Neurath, unlike the other members of the Vienna Circle, approaches philosophical questions from a background in the social sciences rather than in the mathematical exact sciences, and, accordingly, he shows very little interest in either the problem of a priori principles or the problem of elucidating the peculiar position and role of philosophy vis-à-vis the special sciences. See the papers by R. Haller in T. Uebel, ed., *Rediscovering the Forgotten Vienna Circle: Austrian Studies on Otto Neurath and the Vienna Circle* (Dordrecht: Kluwer, 1991), and also T. Uebel, "Neurath on Naturalistic Epistemology," forthcoming.

No real answer to these questions was forthcoming until Carnap's *Logische Syntax der Sprache*[17] of 1934. Here Carnap is once again responding to recent developments in the exact sciences: to Heyting's formalization of intuitionistic arithmetic, for example, and, above all, to Hilbert's program of "metamathematics." Moreover, he is once again attempting to neutralize the philosophical disputes arising in connection with these developments by showing how all parties involved are in possession of *part* of the truth; the remaining part that appears to be in dispute is then argued not to be subject to rational debate at all. More specifically, the dispute is declared to be a matter of convention in precisely Poincaré's sense: there is simply no fact of the matter concerning which party is "correct," and thus the choice between them is merely pragmatic.[18]

The dispute in question arises from increasing appreciation of how fundamentally the program of *Principia Mathematica* is threatened by the paradoxes and involves the three traditional schools in the foundations of mathematics: logicism, formalism, and intuitionism. Carnap responds to this dispute by declaring that each side is simply putting forth a proposal to construct a formal system or calculus of a certain kind: logicism proposes to construct an axiomatization of mathematics using the rules of classical logic, intuitionism proposes to construct an axiomatization using the more restrictive rules of intuitionistic logic, and so on. The essential point is that no such formal system or calculus is more "correct" than any other—indeed, the notion of "correctness" is entirely inappropriate here. Considered simply as proposals to construct formal systems of various kinds all the apparently opposing philosophies are then equally "correct," and the choice between them can only be a purely pragmatic question of convenience. In the "Foreword" to *Logical Syntax,* Carnap calls this standpoint the "principle of tolerance" and then remarks:

> The first attempts to cast the ship of logic off from the *terra firma* of the classical forms were certainly bold ones, considered from the historical point of view. But they were hampered by the striving after "correctness." Now, however, that impediment has been overcome,

---

[17] *The Logical Syntax of Language,* A. Smeaton, trans. (New York: Routledge, 1937)

[18] Cf. again a well-known passage from *Science and Hypothesis* (end of ch. III). "What, then, are we to think of the question: Is Euclidean geometry true? It has no meaning. . . . One geometry cannot be more true than another, it can only be more convenient."

> and before us lies the boundless ocean of unlimited possibilities (*op. cit.*, p. xv).

It is thus with an exuberant sense of liberation that Carnap extends Poincaré's conventionalism to logic itself.

But how is it possible for Carnap to maintain a stance of neutrality with respect to logic itself? Here is where the fundamental insights of Hilbert's metamathematics come into play. For we are to describe the logical rules governing the formal systems or calculi under consideration within the *metadiscipline* of logical syntax: each system is viewed simply as a set of strings of symbols together with rules for manipulating such strings (entirely independently of any question concerning their "meanings"), and any and all formal systems can then be specified from the neutral standpoint of a purely syntactic metalanguage. More precisely, the syntactic metalanguage need employ only the limited resources of primitive recursive arithmetic, and we can thus describe the rules of classical systems, intuitionist systems, and so on from a standpoint that is neutral between them. Our aim is not to justify one system over others as inherently more "correct," but simply to describe the *consequences* of choosing any such system.

Carnap is now in a position precisely to articulate the method and stance of logical analysis. This paradigmatically philosophical enterprise is simply a branch of logical syntax: specifically, the logical syntax of the language of science.[19] We are thus concerned with what Carnap calls "the physical language" (§82). The physical language, as opposed to purely mathematical languages, is characterized by two essentially distinct types of rules: *logical* rules represent the purely formal, nonempirical part of our scientific theory, whereas *physical* rules represent its material or empirical content. Moreover, this purely syntactical distinction between logical and physical rules is Carnap's precise explication for the traditional distinction between analytic and synthetic judgments (§§51–2). Finally, and what most concerns us here, it is also clear that these logical rules or analytic sentences of the language of science represent Carnap's precise explication for the constitutive—but nonsynthetic—notion of aprioricity discussed above: these logical rules, in other words, syntactically represent Poincaré's (and Schlick's) conventions and Reichenbach's axioms of coordination.[20] Accordingly, while logical rules, just as much as physical rules, can indeed be revised in the progress of empirical inquiry, there is still a sharp and fundamental distinction, within the context of any given stage of inquiry, between the two types of rules.

---

[19] This way of conceiving logical analysis is entirely impossible within the conception of logic of *Principia Mathematica*, where there is no distinction between object language and metalanguage and the language of logic is essentially interpreted And it is for this reason that earlier writers such as Schlick and Wittgenstein characterized logical analysis as an activity rather than a doctrine—Carnap here explicitly rejects such "mysticism" (§73). Compare W Goldfarb, "Logic in the Twenties: The Nature of the Quantifier," *The Journal of Symbolic Logic*, XLIV (1979): 351–68, T. Ricketts, "Frege, the *Tractatus*, and the Logocentric Predicament," *Noûs*, XIX (1985)· 3–15

Carnap's solutions to the two problems depicted above as lying at the basis of the logical positivists' radically new philosophical position are therefore as follows: the distinction between conventions or coordinative definitions and empirical laws properly so-called is just the distinction between logical and physical rules, analytic and synthetic sentences; the standpoint and method of philosophy—now conceived as logical analysis—is just the logical syntax of the language of science. Unfortunately, however, it proves to be impossible to implement both of these solutions simultaneously. More precisely, it proves impossible to implement both simultaneously together with what I take to be another linchpin of Carnap's distinctive philosophical stance: the claim to thoroughgoing philosophical neutrality. For it is a consequence of Gödel's incompleteness theorem that, for any language containing classical arithmetic among its logical rules or analytic sentences, the distinction between logical and physical rules can itself be drawn only within a metalanguage essentially richer than classical arithmetic.[21] Implementing Carnap's analytic/synthetic distinction for such a classical language therefore results in a metalanguage that, in particular, is in no way neutral between classical mathematics and intuitionism. It follows that there is no philosophically neutral metaperspective within which Carnap's distinctive version of conventionalism can be coherently articulated, and it is precisely here, it seems to me, that the ultimate failure of logical positivism is to be found.

---

[20] Thus, the first example of §50 parallels Reichenbach's account of the metric of physical space in *The Theory of Relativity and A Priori Knowledge:* it counts as "logical" in the flat space of classical physics but is "descriptive" (empirical) in the variably curved Riemannian space (-time) of general relativity.

[21] For details, see my "Logical Truth and Analyticity in Carnap's *Logical Syntax of Language*," in W. Aspray and P. Kitcher, eds., *History and Philosophy of Modern Mathematics* (Minneapolis: Minnesota UP, 1988). Carnap himself is perfectly clear about the technical situation here, but he fails to recognize its consequences for his claim to philosophical neutrality. Under Tarski's direct influence he later abandoned the definition of analyticity—and indeed the program—of *Logical Syntax*.

Yet the implications of this failure for our contemporary, post-positivist philosophical situation have not, I think, been sufficiently appreciated. What I want to call attention to here are the very substantial parallels between central aspects of our post-positivist situation and basic elements of the positivists' own philosophical position. Thus, for example, it is now clear, I hope, that, far from being naive empiricists, the positivists in fact incorporated what we now call the theory ladenness of observation as central to their novel conception of science—a conception neither strictly empiricist nor strictly Kantian. Accordingly, they also explicitly recognized—and indeed emphasized—types of theoretical change having no straightforwardly rational or factual basis. In Carnap's hands, these conventionalist and pragmatic tendencies even gave rise to a very general version of philosophical "relativism" expressed in the "principle of tolerance." If I am not mistaken, then, Cassirer's well-known characterization of the Romantic reaction against the Enlightenment—that the battle proceeded largely on the basis of weapons forged by the earlier movement itself[22]—is perhaps even more true of the contemporary reaction against logical positivism. Since it proved ultimately impossible to combine all the elements of positivist thought into a single coherent position, it would serve us very well indeed, I suggest, to become as clear as possible about the true character and origins of our own philosophical weapons.

MICHAEL FRIEDMAN

University of Illinois/Chicago

---

[22] E. Cassirer, *The Philosophy of the Enlightenment*, F. Koelln and J. Pettegrove, trans. (Princeton: University Press, 1951), ch. V, p. 197.

LOGICAL POSITIVISM: A RETROSPECTIVE*

POSITIVISM in the nineteenth century was demystifying. It announced, with Comte, the passage to empirical science from theological and metaphysical ways of thinking. Logical positivists shared this antipathy. Metaphysics could only be a corrupting confusion. Cutting it off, like a gangrenous foot, would be an exercise in honorable self-denial. No matter that we limp.

Could it be that positivists were deceived? What if the categorization of all knowledge as analytic and logical, or synthetic and empirical covered over some dubious assumptions at the heart of positivist thinking? What if logical positivism was itself a disguised metaphysics? There are, I suggest, three assumptions of that sort.

First is foundationalism. It is frequently argued that the positivists were *epistemic* foundationalists: they often supposed that knowledge claims should be reconstructed on the basis of assertions that are stipulated or certain. Mathematics was to be derived from an axiomitized logic. Scientific claims were to be reconstructed in ways showing that sensory data provide their only material content. Michael Friedman contests this reading, arguing that some positivists supposed that knowledge claims should be edited and reorganized to express the best available science. Schlick is his example, though Schlick, too, believed that the whole content of any science is supplied by the sensory data organized and represented by its theories. The characterization of sensory data is affected, as Friedman says, by competing views about their proper analysis. A new science, e.g., Gestalt psychology, alters the favored style of analysis. Still, all the material content within a science reduces to sensory data. It is foundational for the conceptual superstructure built upon it. Theories not anchored by the data are speculative, excessive, and metaphysical.

Epistemic foundationalism has its *ontological* correlate. Let Carnap be our example. He supposed that alternative conceptual systems are constructed, then used for thinking about the world. Any thinkable world has just those differentiations and relationships prefigured within the framework of rules used to construct and organize our claims about it. Indeed, the alleged relation between reality and the system used for thinking it is tighter still: reality, so far as it

---

* Abstract of a paper to be presented in an APA symposium on Logical Positivism, December 28, commenting on a paper by Michael Friedman. Richard Creath will also comment. See this JOURNAL, this issue, 505–19 and 522–3, respectively, for their contributions.

0022-362X/91/8810/520–1

is intelligible, is constituted by the rules of the framework used to think it. Every thinkable reality is grounded in the mind that makes it.

Second is Carnap's assumption about the status of this schematizing mind. He supposed that mind is characterized, for the purposes of empirical psychology, in the terms appropriate to whatever framework is used for thinking a world. Yet, this tells us nothing of the thinker who constructs, then chooses among conceptual frameworks. One is an empirical object like the many others characterizable within a semantical framework. The other is that agent whose existence and character are independent of the systems it makes. Carnap assumed what Kant argued: that mind is transcendental.

Third are these questions about the choice of conceptual frameworks: Why is any one constructed in its particular way? Why use one in preference to another for making a thinkable world? The answer is explicit in Carnap's notion of "pragmatics": the mind working transcendentally is motivated by one or another value. What are values themselves? Presumably, just the objectives valorized by transcendental attitudes or desires. Where values drive the invention and choice of conceptual systems, it follows that worlds relativized to conceptual systems are the function of those values. We create such worlds as will satisfy the values held transcendentally.

These three, very contentious assumptions persist in the successors to positivism. Value-driven, world-making minds are assumed, if never described, in thinkers who include Quine, Foucault, Putnam, Derrida, and Rorty. They share the positivist aversion for abductions. These are inferences moving from something specified to its constituents or conditions: we infer from smoke to fire and from the disappearance of light to black holes. Abduction presupposes a reality stretching beyond our sensory data and beyond our conceptualizations, one to which we infer as we explain the sensory affects in us. Carnap declared these inferences external and meaningless. Rorty makes the same point more directly: there is, he says, no "Beyond."

The attitude toward abduction is a litmus test: we are realists prepared to speculate about a world imperfectly known; or we suppose that value-driven mind creates its conceptual systems, and thereby the only thinkable worlds.

DAVID WEISSMAN

City University of New York

CONVENTION, NEUTRALITY, AND THE LIMITS OF LOGIC[*]

THERE has always been a large body of philosophically troublesome claims. They range from logic and Euclidean geometry, on the one hand, to such claims as that temperature is measured with a thermometer, and that observation is trustworthy, on the other. For want of a better word, let us give these claims the technical name "troublesome truths."

The advent of non-Euclidean geometries in the nineteenth century was a blow both to the rationalist faith in our ability to intuit troublesome truths as well as to the Kantian conviction that troublesome truths are imposed by our minds as form upon the content of experience. In addition, empiricism was undermined by the discovery that the same empirical facts could be described in either Euclidean or non-Euclidean ways depending on the choice of measurement procedures.

The logical positivists, as Friedman correctly notes, tried to forge a position midway between Kantianism and empiricism. Even early on, the troublesome truths about measurement procedures were to be conventions, while the axioms of pure geometry were implicit definitions. In *The Logical Syntax of Language*, the serious consideration of nonstandard logics forced Carnap to a vastly generalized version of this move, now expanded to include all troublesome truths. Thus, there is no correct logic, and since philosophy is logic, no correct philosophy. They are all linguistic conventions. Even empiricism was to be a matter of convention as was the form, subject matter, and confidence level of observation reports. While it may be that no language is more true than another, some are more useful than others. Philosophy's task is to explore alternative linguistic conventions in order to reveal their practical consequences. The language of this exploration will, of course, be conventional as well. That is to say that there are alternatives among which none is uniquely correct. Since the metalanguage will itself have some structure or other, strict neutrality in the comparison of object languages will be neither possible nor (contrary to Friedman) necessary. The neutrality that he believes essential is not necessary precisely because no issue of correctness in the choice of an object language is

---

[*] Abstract of a paper to be presented in an APA symposium on Logical Positivism, December 28, commenting on a paper by Michael Friedman. David Weissman will also comment. See this JOURNAL, this issue, 505–19 and 520–1, respectively, for their contributions.

0022-362X/91/8810/522–3

involved. Thus, if you and I come to different practical decisions in the choice of languages, we can perfectly well understand one another and go our separate ways.

Although *Logical Syntax* is one of the most exciting, and far reaching, and little studied philosophical developments from Kant to Quine, it is not without difficulties of its own. First, Carnap's attempt to distinguish logical from descriptive terms fails irreparably. While much of what Carnap says must therefore be revised, his linguistic conventionalism survives, and its philosophic interest is undiminished. Second, his distinction between L-rules and P-rules fails to do what Carnap wants it to, namely, to mark the boundary of the conventional.

Carnap later made various improved attempts to mark this boundary. Whatever the success of these efforts, I think that it is possible for us intelligibly to mark out the boundary of the conventional, even if that boundary is itself conventional and subject to change, and even if it is not uniform across investigators. Obviously, this is not the time and place to try. In any case, it may well be that only by attending to that boundary, by sharpening it, and by self-consciously relativizing what we say to a set of conventions that we can hope for some form of objectivity, albeit a relativized one. Only thus can we come to accept and explore our intellectual and perhaps even political differences, differences that have for too long been either ignored or deprived of rational scrutiny.

RICHARD CREATH

Arizona State University

WHY FICHTE NOW?*

BEGIN with a self-evidently certain starting point or *Grundsatz,* proceed to determine the series of conditions necessary for the possibility of the same, and do not stop until you have thereby "derived" from your initial starting point the first principles of every particular science, as well as the necessary and general features of all experience. Since your original principle posits an act of free self-assertion, all you need to do in order to construct your science is scrupulously observe and describe all of the additional acts that must also be performed by any freely self-positing subject, as conditions for the possibility of its own conscious self-positing. In this way, you will simultaneously have provided a "genetic" account of consciousness and a "transcendental" account of experience, and you will thereby have constructed a systematic science which not only displays the principles governing consciousness of external objects, but which also exhibits the basic principles of natural law and ethics.

Such is the program of J.G. Fichte's *Wissenschaftslehre,* and it would be difficult to imagine a paradigm of philosophical inquiry more completely at odds with those of the late twentieth century. Nevertheless, to the consternation of some and the bewilderment of others, a world-wide revival of interest in Fichte's thought is currently underway,[1] a revival by no means confined to revisionist attempts to reassess the place of the *Wissenschaftslehre* in the history of German Idealism, but giving rise as well to dramatic claims on behalf of the intrinsic merits and contemporary value of Fichte's transcendental philosophy.

---

* To be presented in an APA symposium on Post-Kantian Idealism, December 28. Robert B. Pippin will be co-symposiast, and Chernor M. Jalloh will comment. See this JOURNAL, this issue, 532–41 and 542–3, respectively, for their contributions.

[1] Indications of contemporary interest in Fichte include: the publication of a superb, critical edition of his collected works, *J.G. Fichte: Gesamtausgabe der Bayerischen Akademie der Wissenschaften,* R. Lauth, H. Jacobs, and H. Gliwitzky, eds. [=GA] (Stuttgart-Bad Cannstatt: Frommann, 1964ff.); many new translations of his writings into various European and non-European languages; the establishment, in the last decade, of an International Fichte Society, as well as several national Fichte societies (including a North American Fichte Society); the inauguration (in 1990) of a new journal, *Fichte-Studien,* devoted solely to his philosophy, plus the recent appearance of a number of special "Fichte issues" of several established journals; and, finally, the ever-increasing number of new books and essays devoted to his philosophy.

0022-362X/91/8810/524–31

Without attempting to offer a new interpretation of Fichte's philosophy or to report on the current Fichte renaissance, I here wish to address a question provoked by the latter: Why Fichte now? Or, more specifically, what can philosophers who pride themselves on their "anti-idealism," their "antitranscendentalism," and their "antifoundationalism" possibly hope to gain from an acquaintance with Fichte's *Wissenschaftslehre*? And I wish to answer this question by calling attention to three features of the latter: its account of consciousness as a synthesis of theoretical and practical elements, with a special emphasis upon the "primacy" of the latter; its insistence upon the fundamental and irreducible character of freedom and the implications it draws from this insight; and, finally, the surprisingly sophisticated character of its "foundationalism" and the distinctly "practical" manner in which this commitment to foundationalism is justified.

I

Fichte's account of consciousness is, broadly speaking, "phenomenological," in that it is based upon the evidence of introspection and treats questions of "being" as secondary. Like Descartes, Fichte ascribes a privileged status to the "I think," and, like Kant, he understands this "I think" as tacitly or potentially present in every moment of consciousness. Fichte's own theory of the I goes well beyond anything found in his predecessors, however, and it does so because it explicitly addresses two unresolved problems: (1) How can we explain our thetic consciousness of ourselves without falling into an infinite regress of objects and subjects of consciousness? (2) What is the relationship between our "theoretical" consciousness of objects and our "practical" consciousness of our own willing of a goal, and what is the connection between each of these and the pure "I think" of self-consciousness? (In the language of Kant's critical philosophy: What is the relationship between the cognizing intellect, the categorical imperative, and the transcendental unity of apperception?)

Fichte's theory of subjectivity addresses both of these questions. (1) Infinite regress (of the sort associated with any purely reflective model of consciousness) is avoided by positing an initial, ungrounded act, in which the I simultaneously posits itself, affirms its self-identity, and distinguishes itself from itself. So understood, self-consciousness (and hence consciousness in general) is an *achievement*, an act of self-production as well as self-discovery. (2) The indicated unity of practical and theoretical experience, of the knowing "intellect" and the moral "agent," is established by showing that every moment of consciousness necessarily involves "theoretical" as well as "practical" elements, that is, that object consciousness and goal consciousness are always present together—a thesis sustained by showing how both must be present within the original self-positing of self-consciousness: no consciousness without conscience, and vice versa.

According to Fichte's highly original account of self-identity, the proposition "I am I" is neither a statement of fact nor an empty tautology, but is, instead, an expression of a *demand*: I ought to be in harmony with myself; the object of (theoretical) consciousness *ought* to conform to the will of the (practical) subject. It is in this sense that Fichte can be said to have taken the principle of "the primacy of practical reason" more deeply to heart than did Kant himself. Even theoretical consciousness is comprehensible only as a task, never as a *fait accompli*. By beginning with subjective spontaneity and ending with felt necessity, the Jena *Wissenschaftslehre* attempts "to explain representations with reference to the acting of a free being"; and, insofar as it derives the former from the latter, it not merely *asserts* but also *demonstrates* that "freedom is the ultimate explanatory base of all consciousness."[2]

What particularly distinguishes Fichte's effort to establish the intimate connection between "knowledge" and "interest" from other such efforts is the way in which he attempts to show, within the confines of a general theory of the "equiprimordiality" of theoretical and practical reason, the *primacy* of the latter within any philosophical explanation of experience. What the *Wissenschaftslehre* purports to demonstrate, therefore, is not merely that the dictates of practical reason are more certain than those of theoretical understanding, but also that "our freedom *itself is a theoretical determining principle of our world*" (GA, I, 5, p. 77; cf. WL, p. 61, and GA, I, 2, pp. 65, 399). Hence, Fichte is not content merely to "deny knowledge in order to make room for faith," or to limit theoretical reason in such a way as to preserve a logical space for freedom; he claims to have shown that "the practical I is the I of original self-consciousness, that a rational being perceives itself immediately only in willing, and that it would not perceive itself—and, consequently, would not perceive the world—and would not even be an intellect, if it were not a practical being" (GA, I, 4, p. 332).

II

"My system," claimed Fichte, "is the first system of freedom" and is, "from start to finish, nothing but an analysis of the concept of freedom" (GA, III, 2, p. 298, and III, 4, p. 182). Part of the meaning of this provocative claim has just been indicated, namely, that consciousness of objects cannot be understood without reference to an act of free self-determination on the part of the knowing subject. But note that this initial free (and thus "unconditioned") act, from which everything else is supposed to be "derived," is, at the same time, itself "conditioned" by everything derived as a "condition for *its* possibility." Hence, Fichte's transcendental deduction of the structure of experience also turns out to be a demonstration of the sheer impossibility of absolute, unconditioned freedom. Even though a free agent must consider (that is "posit") her freedom as unlimited in principle, *actual* freedom is always limited and conditioned. Thus, what at first appears to be a defense of "absolute freedom" is, in fact, an argument that real freedom is always *finite* and that real 'I's are necessarily limited, temporal, corporeal creatures.

---

[2] *Wissenschaftslehre nova methodo*, E. Fuchs, ed. [=WL] (Hamburg: Meiner, 1982), p 19 and GA, I, 3, p. 36.

The *Wissenschaftslehre* simultaneously recognizes both the necessary *opposition* between freedom and necessity *and* the necessary *relation* between the two, a relationship best expressed in the concept of "striving." Indeed, what it tries to show is that the only adequate "system of freedom" is the one provided by such as "*Strebungsphilosophie*" (GA, II, 3, p. 265).

There are, however, other aspects of Fichte's discussion of freedom which deserve comment, and pre-eminent among these is the profoundly *social* dimension of the same and the implications this has for an understanding of "human rights" or "natural law." Unlike ethics, the "philosophy of right" is not concerned with what human beings *ought* to do; instead, it concerns itself with the purely transcendental question of how one's own freedom must be constrained in order to recognize the freedom of others—which, according to Fichte's own arguments, one must recognize in order to recognize one's own freedom, and thus in order to be free at all. Hence, Fichte describes the concept of *Recht* as "nothing other than the concept of the *possibility of the coexistence of the freedom of several rational-sensual beings*" (GA, I, 4, p. 299).

This permits a strict separation between questions of morality and questions of justice or "right," and paves the way for a purely transcendental approach to political theory, the full implications of which still remain to be explored (though this has been the focus of valuable recent work by Jean Renaut and Luc Ferry). In any case, Fichte's "system of freedom" has more in common with liberal political theory than is generally realized and offers some unusual and interesting strategies for a theory of human rights.

Another suggestive feature of Fichte's treatment of human freedom is his clear recognition of the extent to which the so-called "debate" over freedom and necessity actually falls *outside* of philosophy and is not really a "debate" at all. Since I have discussed this point elsewhere,[3] here I shall simply reiterate that Fichte believed that a practical conviction that one is a free and responsible moral agent is not something that can be inculcated in anyone by studying philosophy; instead, it is a *precondition* for entry into the *Wissenschaftslehre.*

For all of his grandiose philosophical ambitions, few philosophers have recognized more clearly than Fichte the limits of speculative reason. Even the soundest "system of freedom" cannot prove to you that you are in some sense exempt from the laws of causal necessity, no more than the most consistent materialism could ever prove that you are no more free than "a piece of lava on the moon" (GA, I, 2, p. 326N). This—and this alone—is the meaning of the notorious remark that "the kind of philosophy one chooses depends upon the kind of person one is" (GA, I, 4, p. 195).

III

Finally, let us consider what some might think to be the most backward-looking feature of the Jena *Wissenschaftslehre,* viz., its uncompromising commitment to a deeply foundationalist notion of philosophy's task. Although Fichte was certainly a "foundationalist" in the sense that he conceived of philosophy as an explanatory science, the task of which is to establish the "grounds" or "foundations" of all experience, several unusual features of his foundationalism deserve special mention.

First of all, it is striking indeed that Fichte should always have accompanied his espousal of philosophy's foundational task with a candid recognition of those contextual factors which, in the eyes of others, undermine the very possibility of philosophical foundationalism. Thus, for example, he delighted in calling attention to the unavoidable circularity of all systematic explanations and to the need for even the most "presuppositionless" science to begin by presupposing such things as natural language and the ordinary rules of reflection. For Fichte, however, such admissions do not constitute an argument against foundationalism, but are simply a recognition that philosophy, like every other human endeavor, is an enterprise for finite intellects. Properly understood, such unavoidable factors are not *obstacles* to the foundational task of philosophical explanation, but rather, are *conditions for the possibility of the same.*

---

[3] "How to Make an Idealist," *Philosophical Forum,* XIX, 2–3 (Winter–Spring 1988): 97–123.

This, however, still does not explain why Fichte believed philosophy *ought* to be a foundational science or how he proposed to establish the immediate truth of his own first principle. In fact, his answers to these two questions are very closely related, for the first principle of the *Wissenschaftslehre* contains within itself the demand that practical freedom ought to be reconcilable with natural necessity, i.e., that human experience ought to constitute a single system, within which freedom would serve as the ultimate "ground."

A philosophical first principle, according to Fichte, has to be immediately certain. On the other hand, it was equally clear to him that none of the so-called "facts of consciousness" possesses the requisite type of self-evidence. This realization, however, did not lead Fichte to abandon his project, but rather, to propose a first principle that does not express any "fact" whatsoever. Accordingly, his system begins with an *act* rather than with a *fact,* with something *produced,* rather than with something *discovered.* Moreover, the act with which the *Wissenschaftslehre* commences is a *free* act, and not just any free act; it is an act of practical self-legislation or free self-positing, and thus an act that has freedom as its content as well as its form: an act of positing one's own freedom. The possibility of a "first principle" from which the possibility of experience can be derived (and hence the possibility of philosophy itself as a foundational system) depends upon the unique, self-validating character of such an act.

Accordingly, Fichte responded to the charge that one could always ask questions about the justification and certainty of *any* proposed starting point in philosophy by *acknowledging* the theoretical legitimacy of the criticism, while at the same time offering an essentially *practical* defense both of his claim that there nevertheless must be a first principle and of his proposed first principle itself: "I *cannot* go beyond this standpoint, because I am not *permitted* to do so" (GA, I, 4, p. 219). Here we have an unambiguous statement of the profoundly *practical* (or, if you prefer, moral) character of Fichte's foundationalism. Granted, I could always raise questions about whether I am really free or not; I can always challenge the authority of my own conscience. Nevertheless, I do not do so; instead, I treat my categorical awareness of my own free self-legislation as the stopping point of critical reflection and thus as the starting point of systematic construction. Why do I refuse to raise further questions in this case? "I can go no further," wrote Fichte, "so long, that is, as I do not wish to destroy my own inner self. Therefore the sole reason why I cannot go any further is because I cannot *will* to go any further" (GA, I, 5, p. 351).

Fichte's insistence upon the legitimacy—indeed, the necessity—of his own foundational project is, therefore, ultimately based upon an appeal to the categorical character of one's self-awareness as a free agent, combined with the equally practical demand that our experience ought to possess systematic unity and that our knowledge ought to have a foundation in something certain—or, at least, that one ought to *seek* such a foundation (see GA, I, 2, p. 133). For this reason, his position might be described as "practical" (or even "ethical") foundationalism.

It is interesting to contrast Fichte's frank recognition of this point with what Nietzsche, almost a century later, took to be his "genealogical" exposé of the ultimately *moral* roots of the demand for truth, certainty, and ultimate foundations. What the comparison shows, aside from a perhaps surprising agreement on the central point, is that at least some "foundationalist" philosophers are not merely willing to *concede* Nietzsche's point, but are even anxious to *insist* upon it—indeed, to make it the explicit foundation of their own "foundationalism."

IV

Although I have mentioned but three, there are many other themes from Fichte's writings which seem relevant to some of the concerns of present-day philosophers: for example, his discussion of the "need" for philosophy and, specifically, of the relationship between an adequate philosophical system and the sense of self-alienation or "existential self-contradiction" which produces the need for a philosophical explanation of experience in the first place. Thus he insisted that, despite its unavoidably abstruse and abstract character, the whole *point* of the *Wissenschaftslehre* is to bring about "a complete reconciliation of philosophy and common sense" (GA, III, 2, p. 392; also GA, I, 2, p. 424), and his reflections upon the connection between our everyday, practical striving to be at one with ourselves and our theoretical striving for systematic coherence in our philosophy may still serve as a bracing reminder that philosophy is—or can be—more than an esoteric game.

Also worth reflecting upon is Fichte's contention that any worthwhile philosophy must also be a philosophy of philosophy and hence must address the question of its own possibility. Of particular interest in this regard are his many attempts to explicate the difference between the "ordinary" and the "philosophical" standpoints, or between the "practical" (or "realistic") and the "speculative" (or "idealistic") points of view, as well as his not altogether satisfactory efforts to explain how one can accomplish the "transition" from the one to the other—a question which, as he somewhat ruefully noted, concerns nothing less that "the very possibility of philosophy" (WL, p. 243). However tentative his solutions, few philosophers have been more aware than Fichte that philosophy indeed lays claim to a "view from nowhere," and even fewer have tried harder to justify this claim.

Although I cannot here develop any of these suggestions and hints, I hope that what I have said is sufficient to answer the question "Why Fichte now?" My answer to this question, however, cannot help but raise another one, namely, how viable are Fichte's own proposed *solutions* to the problems he himself raises? How tenable is his philosophy? In order to answer *this* question, one must turn to the *Wissenschaftslehre* itself; and by this I do not mean simply that one must consult Fichte's writings. For, just as he insisted that only those who reach Kant's results in their own way can really be said to understand Kant's philosophy, so he also maintained that "my philosophy should be expounded in an infinite number of different ways. Everyone will have a different way of thinking it—and each person must think of it in a different way in order to think it *at all*" (GA, III, 3, p. 57). Fichte himself, judging by the number of radically dissimilar versions of the *Wissenschaftslehre* he allowed to remain unpublished, never arrived at a presentation of his own system which he considered fully adequate to the task of permitting his own contemporaries to "think the *Wissenschaftslehre*." But if this so, then we are left with yet another, final question: How—if at all—can one "think the *Wissenschaftslehre*" at the close of the twentieth century?

DANIEL BREAZEALE

University of Kentucky

IDEALISM AND AGENCY IN KANT AND HEGEL[*]

KANT and Hegel are well-known idealists. What this means, and what their differences amount to, have obviously been matters for much dispute. Related issues arise when we ask: What are the implications of such idealism and disagreements when we attempt to account for and assess intentional human activity? I discuss such issues in the symposium paper, and summarize the main points in the following.

I

This issue became a famous problem because both concluded that the particular subjective conditions required in order for there to be possible knowledge of nature are, at least prima facie, incompatible with the kind of agency both wanted to defend, namely, a free, or self-determining agency.[1] In Kant, this all led to the necessity for two different "realms" of nature and freedom, and ultimately to a "metaphysics of nature" and an independent "metaphysics of morals." Similarly, a crucial distinction in Hegel's *Encyclopedia* system is between the philosophy of nature and the philosophy of spirit.[2] While Hegel explains this distinction in ways internal to his own system, some of the implications of this division are clear enough. What is important about the "externality" of nature is that nature "exhibits no freedom in existence, but only necessity and contingency."[3] On the other hand, "the essence of spirit" is said to be

---

[*] To be presented in an APA symposium on Post-Kantian Idealism, December 28. Daniel Breazeale will be co-symposiast, and Chernor M. Jalloh will comment. See this JOURNAL, this issue, 524–31 and 542–3, respectively, for their contributions. My thanks to Fred Olafson and Jerry Doppelt for generous comments on an earlier draft.

[1] This ultimately meant for both that I can be a free agent only as a rational agent. Cf. Kant's well-known remarks in the *Foundations of the Metaphysics of Morals*, L. W. Beck, trans. (Indianapolis: Bobbs-Merrill, 1959), p. 29; *Kants gesammelte Schriften* (Berlin: de Gruyter, 1902), Band IV, p. 412; cf. also the *Zusatz* to Paragraph #21 in Hegel's *Philosophy of Right*, T. M. Knox, trans (New York: Oxford, 1952), p. 29; *Grundlinien der Philosophie des Rechts* (Hamburg: Meiner, 1967), pp. 40–1.

[2] Cf. *Hegel's Logic*, W. Wallace, trans. (New York: Oxford, 1975), #18, p. 23; *Enzyklopädie der philosophischen Wissenschaften*, in *Werke in zwanzig Banden* (Frankfurt: Suhrkamp, 1970), Band 8, pp. 63–4. (I have occasionally altered the translation.)

[3] *Hegel's Philosophy of Nature*, A. V. Miller, trans. (New York: Oxford, 1970), #248, p. 17; *Enzyklopädie*, Band 9, p. 27.

0022-362X/91/8810/532–41

"freedom,"[4] and "The substance of spirit is freedom, i.e., the absence of dependence on an other, the relating of self to self."[5]

In both cases, idealism is supposed to help us understand the possibility of two such distinct classifications, or points of view. In Kant's case, the fact that the spatio-temporal and categorial characteristics of nature are subjective conditions rather than metaphysical predicates means that knowledge of nature is knowledge of a phenomenal or appearing world. The noumenal world, or the world considered apart from the finite conditions we require, is unknown. This restriction of knowledge to appearances means that it is at least possible that we are, as noumenal subjects, free, or capable of initiating an action "spontaneously." Kant goes on to make use of this possibility by arguing that it is practically necessary for us to ascribe such agency to ourselves.

Hegel, too, makes use of a claim that knowledge of nature is limited as a way of arguing that the basic assumption involved in claiming a distinct "science of spirit" (freedom) can be asserted without inconsistency. In his case, however, the stress is on the *incompleteness* of the "notions" (*Begriffe*) necessary for an account of nature (or for an account of ourselves as natural beings), rather than on the exclusionary geographical or boundary metaphor made famous in Kant's phenomena/noumena chapter.[6] This suggests that, in his account of the mechanics, physics, and organics of nature, Hegel is trying to show that "Nature is to be regarded as a system of stages, one arising necessarily from the other. . . ."[7] This will mean that, at some stage, we shall need to employ explanatory concepts that require progressively less reference to the strictly material properties of these objects, and more to their functional properties and ultimately to what must be regarded as their own "internal" purposiveness.[8] Of course, similar sorts of claims have been much discussed, and the issues are quite controversial, but for present purposes we need only note that Hegel is a partisan of such antireductive, but antidualist, strategies.

[4] *Hegel's Philosophy of Mind*, W. Wallace, trans. (New York: Oxford, 1971), #382, p. 15; *Enzyklopädie*, Band 10, p. 25.

[5] *Hegel's Philosophy of Mind*, #382, *Zusatz*, p. 15; *Enzyklopädie*, Band 10, #382, Zusatz, p. 26.

[6] *Critique of Pure Reason*, N. Kemp Smith, trans. (New York: St. Martin's, 1929); *Kritik der reinen Vernunft*, in *Kants gesammelte Schriften*, Bds. 3 and 4, A236/B294.

[7] *Hegel's Philosophy of Nature*, #249, p. 20; *Enzyklopädie*, Band 9, p. 31.

[8] *Hegel's Philosophy of Nature*, #376, p. 443, *Enzyklopädie*, Band 9, p. 587.

The traditional picture of the idealism necessary for Hegel to justify such a picture of spirit as a teleological development of nature is as problematic as the traditional dualist Kantian idealism. If Kant's noumenal-phenomenal split implies, as it seems to, a timeless noumenal agency and a psychological determinism in heteronomous action, Hegel's developmentalism appears based on some grand metaphysical claim about the underlying ideal or spiritual nature of reality, first "expressing itself" externally in nature, and then "returning" from such "external" manifestations to the internal, and finally full self-consciousness expressions of "Absolute Spirit." Nature and finite spirit are supposedly compatible because both represent successively incomplete stages or modes of a developing, finally fully self-conscious entity, "World" or "Cosmic Spirit."

Such traditional interpretations have undergone considerable revision in recent years. So-called "double aspect" readings of Kant have made possible an account of agency not committed to the noumenal metaphysics of timeless decisions and have tried to side-step (as irrelevant) the problem of how noumenal decisions can effect changes in the phenomenal world.[9] On such accounts, we understand human actions not by investigating which desires and beliefs caused the action but by understanding the reasoning (implicit or explicit) that justified (or must have justified) the action for the agent. All of this would be compatible with, from a different point of view, still explaining these ratiocinative processes as causally linked episodes in someone's mental history. Agents can thus be said to have an "intelligible" as well as an "empirical" character and there is no inconsistency in their having both because they are not *competing* accounts of the origin of an action.[10]

In a similar spirit, the developmental aspects of Hegel's system should not be understood as a result of some self-moving divine mind, despite Hegel's many metaphors. I have argued that Hegel's general idealist position is essentially a radical extension of Kant's theory of apperceptive subjectivity and of his category theory, and so is not a species of romantic metaphysics.[11] In this context, my argument would suggest that Hegel's primary interest in the *Encyclopedia* concerns the *conceptual* insufficiencies of various aspects or moments of a possible conceptual scheme.[12] Or, even under the assumption that there *are* only natural objects, spatially extended objects in time, at some stage in describing properly and accounting for the complex capacities and functions of such objects, we find that such organisms cannot be adequately explained or even understood by reference to their natural properties.

---

[9] For a summary of this sort of reading, see my *Kant's Theory of Form* (New Haven: Yale, 1982), ch. 7, pp. 188–215.

[10] For the most exhaustive contemporary statement of this position, see H. E. Allison, *Kant's Theory of Freedom* (New York: Cambridge, 1990), and his introductory statement on pp. 5–6. The terms 'intelligible' and 'empirical' character are Kant's.

[11] See my *Hegel's Idealism: The Satisfactions of Self-Consciousness* (New York: Cambridge, 1989).

This returns us to the central dispute already mentioned here. In what sense, at *what* stage, would we want to claim that the natural functions and capacities of an "animal organism" reach a level of complexity such that concepts of nature and naturalistic modes of explanation are insufficient? What kind of insufficiency are we talking about, especially if it is not metaphysical, if it does not require the assumption of some different sort of entity?

Regardless of the answer to this question, the general contrast with Kant's approach is very clear. Hegel's account of the differing conceptual schemes necessary in understanding *Natur* and *Geist* sees these schemes as continuous. Kant, famously, sees them as discontinuous.

II

What sorts of things follow from this disagreement about the special status of the "notions" necessary to account for agency? One of the most famous involves disagreement over moral issues, especially Hegel's charges that Kant's dualism involves a moral theory committed to an "empty formalism" and a "hard-hearted rigorism." With respect to the latter charge, Hegel counters with "the right of particularity," the "right to be satisfied,"[13] and the claim that "moral consciousness cannot forego happiness and leave this element out of its absolute purpose."[14] In other words, the achievement of moral and ethical rightness must not merely be consistent with an individual's happiness and satisfaction. It must be understood as a form of, or "continuous with," such self-satisfaction.

By contrast, Kant's dualistic picture plays out in a familiar way: the necessary conditions for full rational agency require an agent to prescind from the realization of her own substantive good or satisfaction, the kind of end she identifies with, or values, in the real world of sensible desires and contingently formed interests. I could not be said to be *autonomous* in any sort of pursuit of such a substantive end, however "rational" it might otherwise be.

---

[12] See Willem de Vries's useful remarks on supervenience in his *Hegel's Theory of Mental Activity* (Ithaca: Cornell, 1988).

[13] *Philosophy of Right*, #124, *Z*, p. 84; *Grundlinien der Philosophie des Rechts*, p. 112.

[14] *Hegel's Phenomenology of Spirit*, A. V. Miller, trans. (New York: Oxford, 1977), p 366; *Phänomenologie des Geistes* (Hamburg: Meiner, 1952), p. 426.

Hegel argued that this rendered the Kantian ideal of rational agency merely "positive," alien, without any conceivable account of its motivating force in human life. By contrast, Kant had tried to argue explicitly that the realization of rational agency in a moral law does not originally or primarily express some ideal or end (one of the goals we might, or might not, come to identify with or be "motivated" to pursue in the actual world), but for us is experienced always as a constraint or imperative. Whatever general goals *are* required by the moral law (like beneficence) or whatever obligatory ends, moral education, social institutions, or even highest good can be said to be implied by any attempt to obey the law, being a rational agent must mean, originally, a submission to a law, necessarily experienced as an imperative. (Such a submission, of course, need not block or thwart pursuit of my ends, and it need not be unpleasant, but it can only be contingently compatible with such a pursuit or contingently agreeable, since we always remain "radically evil.") For Kant, Hegel's contrasting ideal of a "self-actualization" in his ethics, a full identification or rational reconciliation with my projects, and the ends I come to pursue by my accepting various concrete social and political roles, would involve an unacceptable concession to the mere fact of my sensible desires and ends, and their wholly contingent education and formation by specific historical institutions. This would all be strictly incompatible with the conditions of a fully *self*-determining agency and so human autonomy.

These issues are clearly linked to the "discontinuity/continuity" disagreement in idealism sketched above. As noted, Hegel maintains that if, among the things that are ethically good, the moral good of universally respected rational agency cannot be understood as continuous with all the other things that I, as a particular, natural, historically situated being, have come to desire, care about, or have committed myself to, then there is no way to understand any individual's coming to esteem or value self-determination as such self-legislation, and so no way to understand actually submitting to such a law. Or, if there is something insufficient about the concepts required in a naturalistic (hedonistic, egoistic) understanding of agency, this insufficiency must emerge internally, in unsuccessful attempts to make use of such notions, and whatever resolution emerges will be responsive to, and so continuous with, those specific *aporias.*[15]

More abstractly, Hegel's approach means he is not committed to any sort of libertarian or strictly incompatabilist notion of freedom. I am free if I can get what I want, in those cases where what I want is not imposed on, foreign to, or alien to me, where, regardless of the causal processes by which I came to want it, I can also ultimately come to recognize it as "mine." These are the cases where the goals I pursue can be understood to be rational, and this is a question about the nature of the reasons I, and the entire ethical community, have taken to be authoritative, have come to count as what a free agent would do. It is not a question about the causal power of those reasons.[16]

But this only introduces a number of other issues. To see the contrast between them more clearly, we must first note that Kant's account of rational agency is considerably more complex than the simple dualistic picture suggests. First, as already noted, Kant's crude *diaeresis* of all principles of action into those based on self-love and those based on the moral law, while expressing his general dualistic approach, is *not* meant to parallel strictly his account of our empirical and intelligible character. While it sometimes appears that Kant denies that we can be autonomous in the pursuit of sensible ends because he believes that, in the pursuit of nonmoral ends, we are unfree, moved to action by sensible causes like a desire for pleasure, his considered view does not depend on such determinism and hedonism. He is willing to claim, in *The Doctrine of Virtue,* that "The power to set an end—*any end whatsoever*—is the characteristic of humanity . . . ,"[17] and that in nonmoral action we are responsible and free. But the basis, the information, desires, education, etc., on which we reflect and set ends is simply bound to be limited for finite agents. However rational we are in reflecting on, coordinating, and ranking our various ends, the circumstances of such reflection will always be partial and contingent. This makes it impossible that any such material end could be a reasonable end for *any* rational agent. Moral reasons for acting thus must be qualitatively different from any sort of nonmoral reason.[18]

---

[15] More moderately, Hegel might be taken to mean that reasons in some sense based on "the right thing to do" must be *compatible* with my desire for satisfaction or happiness, that the co-presence of such nonmoral desires as important reasons for my acting, does not taint or qualify the moral worth of the action. See Allen Wood's helpful discussion in *Hegel's Ethical Thought* (New York: Cambridge, 1990), pp. 149, 152–3, 169–73. I think the evidence is that Hegel's own position is stronger, and that he insists on the continuity of ethical reasons with desire-based interests, that the former "realize" the latter, rather than accompany them. The issue is also complicated by Kant's own remarks on the cultivation of morality, virtue, respect, love, and many other topics, which all greatly qualify the standard (and Hegelian) picture of the supreme role of duty in Kantian morality. See Karl Ameriks, "The Hegelian Critique of Kantian Morality," in *New Essays on Kant,* B. den Ouden and M. Moen, eds. (New York: Lang, 1987), pp. 179–212.

[16] Cf. the account in Wood, p. 151.

[17] *The Doctrine of Virtue,* Mary Gregor, trans. (Philadelphia: Pennsylvania UP, 1964), p. 51; *Die Metaphysik der Sitten* in *Kants gesammelte Schriften,* Band 6, p. 391, my emphasis.

Moreover, in recent years, attention to Kant's "*Wille/Willkür*" distinction, and his many remarks about practical rationality in general, have greatly modified the standard picture of the hedonism and psychological determinism supposedly presupposed in Kant's account of nonmoral agency.[19] It is widely claimed now that Kant maintains that I am certainly free and responsible in nonmoral action, that the mere experience of an intensely experienced desire or aversion need not motivate an action unless I "let" it, unless I "elect" to regard that desire *as* a determining ground of action, or take it to be worthy of acting on, "incorporate" it into a maxim.[20]

Ironically, though, this claim that, even in nonmoral action, we must regard ourselves intelligibly, as beings who act on reasons, makes it even *more* difficult for Kant to address himself to Hegel's concerns. This problem again stems from the way Kant understands the rigid distinction between our "intelligible" and "empirical" character, and makes much clearer the real point of Hegel's rigorism charge. Suppose that I am inclined to pursue a policy generally aimed at my own happiness. Suppose that this happiness is not viewed in a crudely hedonistic sense (as aimed at a particular, subjective experience of pleasure), and I am not determined (as a matter of a psychological law) by the mere presence of some inclination, to adopt that policy. In that case, what sorts of considerations, on Kant's view, *could* be relevant to the adoption of such a policy (and could preserve the Kantian conditions of free agency)? There is no good reason to think that something like the quantitative strength of my various inclinations should count as a reason for ranking them and pursuing the strongest. Such a scale, on this reading, could become a principle only if I make it one, and nothing about the strength of such inclinations alone could explain why I would adopt *that* sort of ranking principle. Indeed, if nothing about our empirical character could count as a reason for acting unless already incorporated into a maxim, it is hard to see the possibility of an *original* rationale for any such maxim or policy.

---

[18] Thus to claim, as Kant does, that all nonmoral reasoning is based on the "principle of self-love," need not imply any claim about hedonism. It would mean something like: the formation of ends always occurs from *within* a finite point of view, based on what can be understood and evaluated at a time and place, and so is always formulated with respect to oneself.

[19] This is especially so of passages in *Religion within the Limits of Reason Alone*, T. M. Greene and H. H. Hudson, trans. (New York: Harper, 1960): e.g., ". . . an incentive can determine the will to an action only so far as the individual has incorporated it into his maxim (has made it the general rule in accordance with which he will conduct himself)" (p. 19); *Kants gesammelte Schriften*, Bd. 6, p. 24.

[20] This point is especially stressed throughout Allison's study. See also Andrews Reath, "Hedonism, Heteronomy, and Kant's Principle of Happiness," *Pacific Philosophical Quarterly*, LXX (1989): 42–72.

If the transition in effect, from the psychological experience of interest, aversions, desires, and the like, to counting such inclinations as actually worth acting on, is opaque on the Kantian formulation of a strictly and exclusively intelligible character, what is the alternative? The issues raised here suggest the following.

As we have seen, Hegel accepts the claim that an agent is free, or truly self-determining, only if rational. He disagrees that there is any point in understanding a commitment to a principle of practical rationality as a commitment to "what any rational agent at any place and time" would will. In the first place, Hegel thinks he can show why it is that, at a particular stage in the history of the Christian West, such an "abstractly universal" notion of rationality would have been adopted. The attraction of respecting each other's "purely rational capacities," in abstraction from the divergent and supposedly irreconcilable particular ends and interests of each, is, he claims, comprehensible only within a certain historical narrative. Secondly, he denies there could be any binding moral force for that sort of a criterion, or that practical rationality is even intelligible, given Kant's discontinuous distinction between the empirical and intelligible.

These charges already commit Hegel to two tasks, and both reflect the continuity theme stressed here. First, he must make good on the claim to show that any self-conscious agent, simply *in* attempting to satisfy desires, would be committed to some rationality principle. For Hegel this minimally involves a commitment to a manner of self-satisfaction that could be affirmed by all those whose own self-satisfaction is potentially affected by one's deed. Hegel thus tries to show why a particularly situated, sensible, desiring agent, in attempting to fix objects for those desires, to determine which are worthy of satisfaction, and to order them, would have to come to understand that such a project is itself best fulfilled (or "actualized") in some sort of a universal "rational" system.[21]

More broadly put, Hegel accepts that as "natural beings," the mere presence of the desires, needs, and aversions true of us as a species by virtue of the bodies or biologies we have, all "immediately" count as reasons for actions of various sorts. The need for reflection and choice, however, and the experience of others, *in* the pursuit of these ends alters the context within which they are experienced and creates new sorts of desires and ends. The result of all of this is supposed to be that establishing moral and ethical relations with others does not involve a sacrifice of or abstraction from our particular ends, but their rational "realization."

Secondly, since there are obviously all sorts of ways of "taking others into account" in the formation and satisfaction of one's ends, Hegel must not only show that reflective desire satisfaction commits one in principle to such an ideal, but that the particular liberal-democratic social and political institutions he discusses best realize that ideal. Taking the other into account cannot for Hegel mean merely coordinating my satisfaction with his, or doing nothing morally impermissible in my pursuit of self-interest. It means pursuing the "objects of the will," or the social goals that any modern rational agent would pursue, goals or ends that no reflective agent, seeking satisfaction in a self-determining way, could forego willing.

Or, in a way that oddly parallels many recent reconstructions of Kant influenced by Rawls,[22] Hegel argues that we should indeed treat rational agency as an *ideal* (what any "thinking will," fully self-conscious, would will) and ought to treat institutions that preserve and foster such an end as objects to which rational agents would be committed. But, he claims, if we treat such an ideal as a value, we must be able to show how, under what circumstances, agents pursuing their particular ends could only "realize *themselves*" by not violating and by working to enhance that capacity among all others; that we show that they would be "unrealized" and so irrationally content were they not to pursue such an ideal.[23]

---

[21] *Philosophy of Right*, p. 29; *Grundlinien der Philosophie des Rechts*, p. 39.

[22] I am thinking of various aspects of work by Barbara Herman, Thomas Hill, Jr., Christine Korsgaard, Onora O'Neill, and Andrews Reath, among others, although there are obviously numerous important differences among them, too. See also the helpful discussion in ch. 1 of Terry Pinkard, *Democratic Liberalism and Social Union* (Philadelphia: Temple, 1987), pp. 3–51, and Jerry Doppelt, "Rawls' Kantian Ideal and the Viability of Modern Liberalism," *Inquiry*, XXXI (1989): 413–49.

This, of course, means that whatever socially integrative principles such considerations point to will be the principles appropriate for subjects of specific historical communities. And in that case, trying to justify such principles and practices as more than "locally" rational will shift the general topic of rationality to the rationality of conceptual and thereby (in Hegel at least) cultural change. If we can only determine what counts as "the end any free agent would will" by reference to what counts *to a community* as an essential or rational expression of such agency (e.g., property, voting, choice of spouse or occupation) and what is an incidental or acceptable restriction of freedom or a toleration of chance (market competition, rewards for natural talent), then a full legitimation of such distinctions must include a defense of the historical rationality of such distinctions. In Hegel, this requires some account of the determinate superiority of such principles and practices in comparison with their predecessors and with the options made possible by such predecessors.

At any rate, these are the main tasks of, respectively, the *Philosophy of Right* and the *Phenomenology of Spirit*. Both accounts suggest numerous further issues and problems. I have tried to suggest today a way of understanding how the idealism issue plays out in the famous dispute between Kant and Hegel about morality, and to suggest even more briefly some reasons to reconsider and reassess the nature and value of Hegel's critique of Kant.

ROBERT B PIPPIN

University of California/San Diego

---

[23] This is something that can only be required of agents in, roughly, the post-Enlightenment, modern world. For more on this aspect of the issue, see my *Modernism as a Philosophical Problem: On the Dissatisfactions of European High Culture* (New York: Blackwell, 1991).

FICHTE: FOUNDATIONALISM, ANTIFOUNDATIONALISM, AND THE NEW NIHILISM*

The philosophical issue of foundationalism is the foundation of the modern rebellion against modernism: postmodernism. From Michel Foucault to Jacques Derrida, the shrill battle cry of postmodernistic discontent is the same: antifoundationalism. Supposedly, Nietzsche is the source of this antifoundationalism.

Query: Is antifoundationalism pre-Nietzchean?

I have argued elsewhere that Fichte's *Grundlage der gesamten Wissenschaftslehre* is not a "Wissenschaftslehre." This thesis implies that Fichte is not a foundationalist. Moreover, I have argued that precisely because no absolute I is, nor can be, given in the *Grundlage* (a thesis that Breazeale seems to posit), Hegel's interpretation and critique of Fichte and the tradition founded thereon are completely illusory.

In my opinion, the *Grundlage* can best be read as a systematic demonstration of the impossibility of any foundationalist program. It is ironic, therefore, that the hegemonic Hegelian reading of Fichte has succeeded in convincing most of those who have the time to do battle with Fichte's arcane texts that Fichte is a bumbling foundationalist.

Hegel was fundamentally unjust to Fichte. This injustice has to be cured. But how? I have argued that no Fichte scholar can refute Hegel's (mis)interpretation of Fichte by uncritically accepting Hegel's categorization of the *Grundlage* as a "Wissenschaftslehre." Here is the brilliance of the Hegelian ruse: hang Fichte with his own rope—*Über den Begriff der Wissenschaftslehre*. The argument takes the following shape. If *Über den Begriff der Wissenschaftslehre* is right, then the *Grundlage* is not a "Wissenschaftslehre"; on the other hand, if the *Grundlage* is a "Wissenschaftslehre," then *Über den Begriff der Wissenschaftslehre* must have been written by a fool.

In truth, however, these strategic moves are illusory: the "foundationalism" suggested in *Über den Begriff der Wissenschaftslehre* is shown to be impossible in the *Grundlage*. There is no way anybody

---

* Abstract of a paper to be presented in an APA Symposium on Post-Kantian Idealism, December 28, commenting on papers by Daniel Breazeale and Robert B. Pippin. See this JOURNAL, this issue, 524–31 and 532–41, respectively, for their contributions

0022-362X/91/8810/542–3

can reasonably "reduce" the "I as an Idea" to a foundation, to something that stands, for it is the horizon that is always coming. This is the new opening in Fichte research: Fichte *contra* Nietzsche. Yet it is a dangerous opening.

Even if we were to agree that Fichte was not a foundationalist, it does not follow that Fichte can now be posited *à corps perdu* in the pantheon of postmodernism. I predict, however, that this will be the next unjust act of some enthusiastic philosopher.

To begin with, postmodernism does not believe in reason. Fichte did. Secondly, and this demonstrates the courage of Fichte, postmodernism has yet to muster the courage to demonstrate the systematic *indeterminacy* of antireason. Fichte's *Grundlage* is the demonstration of the systematic indeterminacy of reason. Third, Fichte was no dreamer; man is stuck in indeterminacy and incompleteness. No announcement of the "death of man" here; no exit. Rather, Fichte constantly reminds his reader of the complete "circularity" of human thought, and this vicious thought—the circle that circles itself and is thus based on nothing—is Fichte's revenge on a deconstructed, post-Hegelian *Weltanschauung*.

The issue, therefore, is nihilism. What follows from the unavoidable circularity of philosophy? Decisionism? But is not decisionism but another word for the content of the concept of nihilism?

When we apply Fichte to the multiculturalist rumblings in the academy, we fly back in astonishment. Is the Western canon grounded in a circular, unjustified, exclusionary striving? Even before we could lay down an answer to this question, if a question it be, another Fichtean problematic asserts itself: To the extent that cultures, or ideations thereof, are circular, what would be the content of a successful multiculturalist Western regime?

Nietzsche's dream of an "after-man" can now be answered. To the extent that the beautiful "being" that succeeds "man" thinks at all, such a "being" still has to learn a thing or two from Fichte. That learning, however, will not liberate the will from the necessities of finitude. Rather, it will moderate the will. But this returns us to Fichte's battle cry: finitude is a necessary contradiction in the "human spirit."

CHERNOR M. JALLOH

Howard University

IS MUSIC AN ART?*

> "Goethe ridicules the grouping together of all the arts which are so different from each other in their aims and means of expression. . . ."
>
> Paul Oskar Kristeller

Since about the middle of the eighteenth century, but not before, the term 'Art'—Art with a capital 'A'—has been taken to comprise "above all the five major arts of painting, sculpture, architecture, music, and poetry." As Kristeller,[1] discoverer and principal expositor of this important historical event, observes: "These five [now] constitute the irreducible members of the modern system of the arts, on which all writers and thinkers seem to agree" (*ibid.*, pp. 164–5). In light of Kristeller's discovery, my thesis in the present paper can be very concisely put. It is simply that music is *not* one of the Arts with a capital 'A', which is to say, *not* one of the fine arts.

I

Now, it might be objected straightaway that this is, on the very face of it, an utterly absurd, completely indefensible thesis, since music just is, as a matter of fact, now classified as one of the Arts, one of the fine arts, and has been since 1746, to be exact, so there's an end on't. Whatever is so classified is just that very thing, just as whatever is sold in a grocery store is groceries, regardless of what philosophical analysis or marketing research might reveal.

Well, of course, I am not denying that music is now classified as a fine art. But that it is is a fairly recent historical event. And we customarily represent the event not as *constituting* music Art from then on but as drawing our attention to the fact—getting us to see that—music was Art all along, at least as far back as one wishes to push the repertory of Western art music. But if it makes sense to claim that before the eighteenth century people did not know that what they called music was Art, and mistakenly thought it was something else, then it would seem also to make sense at least to suggest that, actually, they may have been right all along and the eighteenth-century philosophers mistaken in reclassifying music as one of the

---

* To be presented in an APA symposium on Is Music An Art? December 29. Philip Alperson will comment. See this JOURNAL, this issue, 555–6, for his contribution.

[1] "The Modern System of the Arts," repr. in Kristeller, *Renaissance Thought and the Arts: Collected Essays* (Princeton: University Press, 1980).

0022-362X/91/8810/544–54

fine arts. The claim, though perhaps dead wrong, is not unintelligible or silly.

II

My thesis, then, is that music is not Art. But things are not quite so simply put. Specifically, what I am saying is that pure instrumental music, music without text, title, program, or plot—what the nineteenth century came to call "absolute music"—is not Art. Further, I am going to claim that Renaissance polyphony (I shall not push my argument further back than that) is also best understood, prior to the middle of the sixteenth century, as not one of the fine arts. And this claim is meant to include not merely what was then the more or less minor genre of pure instrumental music, but, more importantly, the musical setting of sacred texts as well. Finally, I am going to claim that, beginning in the second half of the sixteenth century, the musical setting of texts—which is agreed on all hands to have undergone a profound change, under the influence of the Counter Reformation, the Council of Trent (1545–1563), and, at the end of the century, the invention of opera—presents the clearest case for a kind of music that *is* Art.

So I am arguing that prior to the middle of the sixteenth century, music is best understood not as a fine art but as something else, to be specified shortly. After that, it undergoes a profound change that encourages thinking of it as Art. And, ironically, in the eighteenth century, just when it is finally swallowed up into the system of the fine arts, it gives birth to pure instrumental music in its modern, imposing form: just the very music that immediately comes to mind when the art of music is mentioned—and just the very music that has been most difficult to place in a theoretical framework capable of enfranchising *it* as Art, while at the same time enfranchising the visual arts and literature as well.

The solution to this dilemma lies simply in recognizing that the eighteenth century passed on to us a problem intractable because based on a false premise. More exactly, the modern system of the arts was formed before absolute music in its later, imposing form was an issue. What the eighteenth-century theorists had in mind, in these formative years, when they baptized "music" Art was always, essentially, music that accompanied a text. And as we shall see, the baptism was not implausible, given that absolute music was barely a consideration at all. And when it became an issue, as the age of Haydn, Mozart, and Beethoven rolled around, the die was already cast. Music was now in the modern system—Art—and absolute music must be gathered in, recalcitrant though it was. It is time now to correct that original false step, by recapturing for absolute music the pre-Tridentine concept of what music is, which the eighteenth century never completely gave up, where absolute music was concerned.

III

What place, then, did music occupy, among human institutions, in the Renaissance, before the momentous events that so transformed it at the end of the sixteenth century? If we confine ourselves to composing, performing, and the experience of listening to music, to the exclusion of music theory, it seems obvious enough to be a truism that music was seen as a craft. But what *kind* of a craft?

In answering this question, let me appeal not to historical texts but, rather, to the contemporary musical ear. Listen, for example, to the superb mass settings of the great fifteenth-century master, Johannes Ockeghem; and compare them to the more familiar polyphony of Palestrina a century later. What must strike one immediately in the comparison is a kind of remoteness in the earlier composer—not so much an other-worldly quality but something, it seems to me, removed in some way from the "human," if I may so put it. Worldly, then, but not human.

The comparison, pressed a bit further, can, I think, fix this impression more clearly. What is "human" about Palestrina? To my ear it is the quality of being a human "utterance"; and in a moment we will have some historical backing for this claim. Palestrina's text sounds as if it is being declaimed, enunciated by a persona, or, better, personae whose intention it is to convey the meaning of the text. I get no such impression from Ockeghem's sacred works, which seem to me, rather, not to be utterances at all but opulent decorative structures: not expressions of the sacred words but *settings* of them in the very sense in which a goldsmith might make a setting of a precious gem.[2] (One wonders, indeed, if our very idiom for describing what composers do with texts, namely, that they "set" them, might not be a vestigial reminder of this earlier compositional attitude.)

But if my impression is correct, then we can now say just what kind of craft music was prior to the middle of the sixteenth century. It was akin to the jeweler's craft; or, as we might call it now, being in possession of the modern system of the arts, one of the *decorative* arts, to distinguish it from Art. The composer was a jeweler, and the sacred texts his precious gems, to be placed in ever more elaborate and dazzling settings, until the words became dwarfed by their frames, like relics in their golden and crystal reliquaries. This is not to say that the words became disvalued. Quite to the contrary, the overpowering, complex beauty of their musical settings is a direct and ample expression of the value beyond price that was placed on them by the composers and their world.

---

[2] I have expressed this thought previously in *Osmin's Rage: Philosophical Reflections on Opera, Drama and Text* (Princeton: University Press, 1988), ch I.

Sacred polyphony, in the first half of the fifteenth century, displays, I think, just because it was a supremely decorative art, an exuberant freedom of the musical imagination that is still impressive today, even to an age familiar with the concept and reality of absolute music. But this freedom entailed, as the spokesmen for the Counter Reformation and the Council of Trent made abundantly clear, a sacrifice to the intelligibility of the text that finally became unacceptable. It is a more than twice-told tale that polyphony was in jeopardy at the Council of Trent just because it was seen as the villain of the piece. How can you understand the message when everyone is saying something different, simultaneously? But sacred polyphony survived, Palestrina its supreme master in the post-Tridentine period, with the imprimatur of the Church. And this in itself should suggest that polyphony was not really the perpetrator at all; rather, as it turns out, merely an accessory before the fact. As Don Harrán,[3] an eminent authority on the relation between tone and text in the early history of Western music observes: "Composers contemporary with and preceding Josquin may at times have been too preoccupied with technical considerations, particularly those attendant upon a polyphonic coordination of voices, to pay attention to individual word stress" (*ibid.* p. 335). Thus, apparently, it was not polyphony per se that obscured the text, but the pre-Tridentine polyphonists' preoccupation with the musical setting itself as pure decorative or structural art. It was not the many-voiced structure that obscured the text but what the individual voices were doing, namely, weaving long, sinuous, melismatic threads that spread the syllables so far apart as to make them unintelligible as words and sentences. To quote the same eminent authority once again, "the recommendations of the Council of Trent . . . were engendered by the concern of the ecclesiastics lest music prevail over words and prevent their understanding"; and the leading recommendation of all to composers was "to curtail melody for the sake of verbal clarification" (*ibid.*, pp. 345–6).

---

[3] *Word-Tone Relations in Musical Thought: From Antiquity to the Seventeenth Century* (Stuttgart: Hänssler, 1986).

To say that composers of Catholic sacred music in obedience to the directives of the Council of Trent *curtailed* melody for the sake of verbal clarification, though true, is to place what they did in an unnecessarily negative light. Melody was indeed curtailed; but, as I view what happened, it was curtailed not in the sense of sacrificed in obedience to authority, contrary to any aesthetic inclination whatever, but, rather, given a different, certainly less flamboyant character, in pursuit of a new, *positive* aesthetic program aimed at achieving the desired objective of textual intelligibility. And, again, I think we can make this clear with a comparison, this time between the polyphony of Palestrina and what can be looked upon as the culmination of the movement toward textual intelligibility in the invention at the end of the century of operatic declamation: the so-called *stile rappresentativo*. For it is operatic monody, both in theory and in practice, that makes unmistakably clear what has been going on. Music has been transformed from the decorative craft of text setting into the fine art of utterance representation. The first composers of opera, as one of their spokesmen, Pietro de'Bardi,[4] put it, "found a way of imitating familiar speech . . ." (*ibid.*, p. 364).

Thus, it was through the positive aesthetic program of artistically representing human utterance, not the negative one of simply trimming their melodic sails, that the inventors of opera achieved, to whatever extent they did, the intelligibility of the text. And in seeing their explicitly stated program, as well as their musical results as the culmination of a historical process that was already in progress during the Counter Reformation and immediately after the Council of Trent, it becomes abundantly clear that such representation of human utterance was already in place in post-Tridentine sacred polyphony—which is precisely why Palestrina possesses that "human" quality of expression that the music of Ockeghem does not. The stage was now set for music to be initiated into the modern system of the arts. For it now was squarely within the category of the representational, which of course would eventually be the defining property of the fine arts when they received their capital 'A' in the eighteenth century. For vocal music it was already, essentially, a *fait accompli* at the end of the sixteenth.

IV

"The decisive step toward a system of the fine arts was taken by the Abbé Batteux," Kristeller tells us, who "was the first to set forth a clearcut system of the fine arts in a treatise devoted exclusively to this subject" (*op. cit.*, p. 199). Batteux' leading idea—his ***même principe,***[5] as the title of his work calls it—was, of course, the Platonic and Aristotelian concept of "imitation." And it remained the leading idea of almost all writers throughout the eighteenth century. As Batteux put the momentous conclusion: "We can define Painting, Sculpture, Dance, as an imitation of beautiful Nature, expressed by colors, by relief, by attitudes. And Music and Poetry an imitation of beautiful Nature expressed by sounds, or by measured discourse" (*ibid.*, p. 43).

---

[4] "Letter to G B. Doni," in Oliver Strunk, *Source Readings in Music History: From Classical Antiquity through the Romantic Era* (New York: Norton, 1950).

Music, as we have seen, was ready for this development, and had been since the late sixteenth century. For it already was by that time, both by practice and by precept, an "imitation of beautiful nature" in the form of a representation of human vocal utterance. And it is as the representation of human utterance, and other forms and mechanisms of human emotive expression in general that the ***même principe*** in Batteux and the rest was almost invariably understood, where music was concerned, throughout the eighteenth century.

But it is absolutely crucial to remember—I have been guilty of forgetting it myself, at times—that what was being talked about almost exclusively during the formative years of the modern system of the arts when music was the subject was ***music with a sung text,*** most particularly the controversial arts of opera and music drama. And it is not at all clear that pure instrumental music even as late as Kant's *Critique of Judgment*—which is to say, the 1790s—was yet ***firmly*** in place as Art.

The classification of pure instrumental music, it is also important to note, was inextricably entangled in the closely related activity of ***evaluating*** it, throughout the eighteenth century and well into the nineteenth; for in classifying absolute music as Art, one was, in effect, elevating it as well. So that it is fairly safe, at least in the eighteenth century, to read the low esteem in which philosophers and critics tended to hold pure instrumental music, even as late as mature Haydn and Mozart, as, ipso facto, a reluctance as well to classify it with the fine arts. And that almost universally held evaluation is a familiar and well-established fact. As David P. Schroeder,[6] a recent commentator on this aspect of the musical scene, concludes: "Until near the end of the eighteenth century, the overwhelming majority of critics held the opinion that instrumental music was vastly inferior to vocal music, and for many the reason lay in the apparent inability of instrumental music to convey moral instruction or, indeed, to attain any level of intelligibility" (*ibid.*, p. 64). Indeed, even as late as Hegel, music without a text still had, in spite of the towering figure of Beethoven behind it, an only grudging acceptance on the part of many, including Hegel[7] himself, who insisted that "music must, on account of its one-sidedness, call on the help of the more exact meaning of words and, in order to become more firmly conjoined with the detail and characteristic expression of the subject matter, it demands a text which alone gives a content to the subjective life's outpouring in the notes" (*ibid.*, p. 960).

---

[5] *Les beaux arts reduits a un **même** principe* (Paris, 1747).

[6] *Haydn and the Enlightenment: The Late Symphonies and their Audience* (New York: Oxford, 1990).

Kant[8] had, indeed, nominally classified pure music as one of the fine arts. But he, in the end, held open the distinct possibility that, although pure musical pleasure might be "the effect of an estimate of form in the play of a number of sensations," it might also be a pleasure afforded by "mere sense-impressions"; and he concluded: "According to the former interpretation, alone, would music be represented out and out as a *fine art,* whereas according to the latter it would be represented as (in part at least) an *agreeable* art" (*ibid.*, § 51, p. 190). Thus, in the work that is agreed on all hands to represent the culmination of aesthetic thought in the eighteenth century—the most profound contribution up to that time to the philosophy of art as we understand it—the issue of absolute music as Art was still in the balance.

The balance was only tipped decisively, for absolute music, in the direction of the fine arts in the Romantic era. And it is of particular significance to my argument that this is the very period in which instrumental music underwent the greatest pressure in its history in the direction of the literary and pictorial arts, eventuating in the development of the tone poem, programmatic symphony, and various other forms of illustrational music: surely a clear case of Mohammed coming to the mountain. The failure of theory to produce the desired result of conceptually connecting absolute music with the firmly established and clearly representational fine arts of literature, painting, and sculpture led instead to an alteration in musical practice itself, thus producing a plethora of "literary" and

---

[7] *Aesthetics: Lectures on Fine Art,* T. M. Knox, trans. (New York: Oxford, 1975), vol. II.

[8] *Critique of Aesthetic Judgment,* James Creed Meredith, trans. (New York: Oxford, 1911).

"painterly" works for instruments alone that could readily be seen as of a kind with poems, novels, plays, paintings, and statues. But, of course, what they could *not* any longer be seen as was works of absolute music. And the reverse side of this coin was, deplorably, an increased tendency to put literary or painterly interpretations on the pure musical canon. The strategy then consisted in both transforming present musical practice to accord with theory and misinterpreting past musical practice to the same end. The result for the philosophy of art was no progress at all with the question.

Here endeth my historical tale. What can we learn from it?

V

Philosophers of art, for a long time, have accepted it as an established historical fact that "music," *sans phrase,* came to be seen as one of the fine arts in the early eighteenth century, with Batteux finally, but *merely* dotting the 'i' in 1746. In addition, we have come to realize that aesthetics and the philosophy of art also took on their modern, philosophically autonomous character during the very same crucial period; and so, like epistemologists, philosophers of mind, philosophical psychologists, and various others, philosophers of art tend to see the sources, primal formulations, and initial solutions of their problems as coming in the early modern period. The heavy hand of a nearly three-hundred-year history, then, as well as the whole philosophical weight of the Enlightenment, have laid on aestheticians the firm conviction that the membership of music, including absolute music, in the modern system of the arts is an established, long-standing, and incontrovertible fact; and it has laid upon them also, in consequence, what they perceive as the unavoidable task of finding, where three centuries of predecessors have failed, a theory to knit absolute music conceptually to the rest of the fine arts. The task has proved intractable. The question then forces itself upon us: Is the task really unavoidable?

One thing we can certainly accomplish by a fresh and unbiased look at the history of musical thought and practice is to lift that heavy historical hand and philosophical weight. It simply is not the case that absolute music ever became firmly established in the eighteenth century as one of the fine arts. The question was not settled then, even as late as Kant's *Critique of Judgment.* The firm historical tradition is far shorter and more problematical than we have tended to think. And if you are intimidated by the authority of history, and the weight of Enlightenment thought from Shaftesbury through Kant, ease your mind. Neither is nearly as imposing as we have been led to believe.

I have come firmly to believe that the history of thinking about the nature of music as an art, and the practice of music itself, from the pre-Tridentine Renaissance through the *Critique of Judgment* has an important philosophical truth to tell us, or, at least, an important philosophical hypothesis to propose, worth considering seriously. I have come to this truth, or this hypothesis (if you prefer), on independent, ahistorical grounds, in my thinking about absolute music; but I am convinced that the history of musical thought lends it support. Prior to the Counter Reformation, music was generally perceived as among the crafts, and in its primary function, as the setting of religious texts, seems rather analogous to the particular craft of setting precious jewels. It is, in short, what we would call now one of the decorative arts. Various intellectual movements in the second half of the sixteenth century, notably, the Counter Reformation and the rise of opera, transformed the musical setting of texts—still, and for a long time to come, the major work of composers—into a palpably representational art, and recognized as such, that is, as *representational,* although not yet as Art, since, as Kristeller observes, "the Renaissance . . . was still far from establishing the modern system of the arts . . ." (*op. cit.,* p. 186).

In the eighteenth century, the musical setting of texts—which usually still was all that "music" meant—was emphatically recognized as one of the fine (that is imitative or representational) arts. But pure instrumental music never, as I now read the historical record, was firmly established as one of the Arts, for the obvious reason that it was, and remains impossible to construe as "representational" in any full-blooded sense of that concept. And Kant was not alone in his century in harboring doubts, recognizing the distinct possibility that the art of absolute music might well be more akin to the decorative arts than to the arts of the modern system.

I suggest that philosophical thinkers about music until the end of the eighteenth century had things exactly right, and that the nineteenth century began to get them exactly wrong. Since the end of the sixteenth century there have been *two* arts of music: the fine art of musical text setting, which is basically the art of representing human expression in musical tones, and the decorative art of absolute music. Recognizing this simple fact frees us from the impossible philosophical task of finding a theory or some other kind of conceptual analysis that can enable us to understand absolute music as one of the fine arts. In a word, it *is not.*

But while liberating us from a pseudo problem we should never have become involved with in the first place, the recognition of absolute music as a decorative rather than a fine art opens up to us a host of new and, to me, both exciting and important problems that may have implications quite beyond the confines of musical aesthetics. These problems of which I speak are all special cases or implications of what might be called the "prestige" question.

Absolute music until the end of the eighteenth century was, as we have seen, far from the center of attention, the setting of religious and dramatic texts being the main business of composers. It was held in low esteem, compared with vocal music, well into the nineteenth century, particularly among philosophers of art. Nevertheless, in spite of reluctance born, clearly, of a pronounced bias toward the content-laden arts, absolute music did finally achieve an exalted place in the Pantheon. That place is now an incontrovertible fact of the art world. And the mere suggestion that absolute music be understood as a decorative rather than a fine art is bound to be seen as a demotion, an affront to the dignity and intellectual stature of the discipline: indeed, an impossible suggestion, a nonstarter just because it cannot account for the prestige pure instrumental music now enjoys in our intellectual lives.

But I would suggest that such a response is both an evasion of the issue, and at the same time expresses a deeply held Western prejudice in favor of the arts of representation which absolutely demands examination. I say it is an evasion of the issue just because as philosophers of art we have failed to take seriously the possibility that absolute music is as it appears, an art of pure sonic design which, for that very reason, has called forth such terms as "absolute" and "pure" to describe it, and to distinguish it from other musical kinds. Instead of accepting this, at least tentatively, as a working hypothesis, so that we can explore its implications and possible foundations, we have persistently applied our philosophical ingenuity to the task of transforming absolute music into one kind or another of crypto-representation, on the sometimes explicit but all too frequently unexpressed premise that only as representation can it really be taken seriously, which, manifestly, it is.

But the question we should be asking is not: "How does absolute music, appearances to the contrary, manage to represent?" That is the evasion. Rather, we should be asking: "What is it about *us*, and about our world, that has made the pure contentless art of musical design so important for our lives in the past one-hundred-and-fifty years, but not before? What needs of ours does it serve? And why have other peoples not felt those needs, or we until so recently in our history? What distinguishes sonic from visual design, that seems to make the former play so different a role in our lives from the latter? What does it all *mean*—but not in the semantic sense?

I have only just begun to address such questions in my own work;[9] and I urge other philosophers of art to take up the task as well. It is a hard task, because we have paid little attention to the decorative arts in our philosophies, tend to disvalue them in comparison to the arts of content, and are, at present, ill-equipped to talk about them.

It will be important, I think, to look to others for help: to the cognitive psychologists, perhaps; the philosophers of mind, structural linguists, and to others, like E. H. Gombrich,[10] who has written so brilliantly about the decorative arts of visual design and who, by the way, quite naturally ends his book on the decorative arts with a chapter on *music*.

I think this is a great challenge, and will spare us another dreary round of philosophical books and papers trying to understand absolute music as a representational or linguistic art. The books and papers get more and more clever, more and more philosophically sophisticated, and, for me at least, less and less plausible. The lady doth protest too much, methinks. I believe that, until we meet this challenge head on, and try philosophically to understand pure music on its own terms—which is to say as a pure decorative art—we will remain with an important and, to date, mysterious aspect of our artistic natures unexplained and, worse still, unexamined.

PETER KIVY

Rutgers University

---

[9] "The Fine Art of Repetition" (work in progress).

[10] *The Sense of Order: A Study in the Psychology of Decorative Art* (Ithaca: Cornell, 1979), Epilogue: *Some Musical Analogues*.

THE ARTS OF MUSIC*

In "Is Music an Art?" Peter Kivy answers his own title question with the intriguing and persuasively argued suggestion that pure instrumental (or "absolute") music is not an art in the sense of that term most likely to be applied to it in the present day, that is, as a fine art. That use of the term, Kivy argues, was enshrined in the eighteenth century and its application to music rests on the notion of an essentially representational practice. It was vocal music of a certain sort, however—the musical declamatory "utterances" of composers such as Palestrina—which made the inclusion of music in the modern system of the arts at all possible, Kivy argues. Subsequent theorists have struggled with the implications of extending that courtesy to instrumental music. But pure instrumental music, Kivy argues, is actually closer in spirit to the pre-Tridentine decorative craft of, say, Ockeghem, than to the music of the late Renaissance. Once we recognize that instrumental music is a decorative and not a fine art, Kivy argues, we will cease to be bedeviled by pictorial and linguistic accounts of instrumental music and we may profitably turn our attention to the value of the art of pure sonic design on its own terms.

In my paper, I push the implications of Kivy's excellent essay in two directions. First, I explore the question whether instrumental music can properly be considered a pure decorative art. I offer an account of the nature and aesthetic value of the arts of decoration generally and consider the senses in which instrumental music might be brought under that rubric. This involves an analysis of the varieties of musical formalism and an estimation of the contribution of such an art to culture.

Second, I complicate Kivy's account of the notion of a "fine art" and of the "modern system of the arts" by employing and pressing Francis Sparshott's distinction between the fine arts considered as arts of imitative play and the fine arts considered as arts of beauty. I do not here dispute the account of the development of the modern system of the arts in the eighteenth century advanced by Kivy, Paul O. Kristeller, and others, but argue, rather, that, even if the theo-

---

* Abstract of a paper to be read in an APA symposium on Is Music An Art? December 29, commenting on a paper by Peter Kivy. See this JOURNAL, this issue, 544–54, for his contribution.

0022-362X/91/8810/555–6

rists in question did not have instrumental music in mind, they well could have. I argue that there are, in fact, at least two "fine arts" of instrumental music (one of which accommodates the understanding of instrumental music as a decorative art) as well as other arts of instrumental music. I examine the sorts of appreciation appropriate to several such arts and the underlying values of these musical practices. I conclude with some general observations about the implications of adopting a pluralistic view of the arts of music.

PHILIP ALPERSON

University of Louisville

PHILOSOPHY IN A DIFFERENT VOICE*

We tell our students philosophy will do valuable things for them: teach them how to think and express themselves clearly, make them more critical of customary beliefs and assumptions, give them a general framework with which to help them make sense of their experiences, give them guidance in everyday life, and the like. But I wonder whether philosophy delivers on these promises.

I shall concern myself with the situation of women students in particular, though at least some of what I have to say will have application to men of color and, ultimately, all men as well. The question I want to raise is: What can our women students get out of their philosophy courses? Now surely, you may say, this is a very strange question to raise, for surely our women students get out of their philosophy courses whatever our men students get, and surely, exactly what that is varies with the person, the teacher, the course, and so forth. But questions like the one posed have proven very significant in other academic fields, for example, in history and English literature. Thus: What can women students get out of their history courses, when the questions and concepts and categories and methods those courses are concerned with are so profoundly geared to men's experiences, leaving women almost wholly invisible? And: What can women students get out of their English literature courses, when so few women writers appear on the course reading lists and when so many of the works that do appear on those lists present women in sexually stereotyped ways—as "whores, bitches, muses, and heroines dead in childbirth," to quote Annette Kolodny. In the cases of history and English literature, women students can get something very different from men students out of their courses, something a lot less relevant to their lives and something a lot less worthwhile.

So what about philosophy, which has traditionally been thought to be concerned with the fundamental questions of human existence in a universal, timeless, and gender-neutral way, to be concerned with such issues as the nature and justification of morality, the nature of truth and beauty, the nature of human knowledge and its justifica-

---

* To be presented in an APA symposium on Philosophy in a Different Voice, December 29. Pamela Hall will comment. See this JOURNAL, this issue, 568–9, for her contribution. I would like to thank Charlene Avallone of Notre Dame's Gender Studies Program for her valuable comments on an earlier version of this paper.

0022-362X/91/8810/557–67

tion, and so forth. What can our women students get out of their philosophy courses? Scholars now working in the area assure us that there have been women philosophers throughout the history of Western civilization and probably in other cultures as well, representing every subspecialty of philosophy save for the history of philosophy (at least prior to this century).[1] Almost all of these women philosophers were part of larger intellectual circles that included other philosophers and learned persons, a few had large followings, and some ancient ones even directed or co-directed schools of philosophy. Some were royalty or aristocrats or had wealthy patrons, though some were self-supporting. Most were trained by their male and female predecessors, though some were self-taught, some educated by their parents, some trained in academies, and some trained in monasteries, usually by other women. And, like men, they wrote in a variety of literary forms, published and unpublished, including formal essays, books, and articles, as well as dialogues, poetry, epic novels, letters, and diaries. All of these women philosophers contributed to the development of philosophic thought. But with very few exceptions, all very recent, women philosophers do not appear on students' reading lists, or in their textbooks, anthologies, histories of philosophy, or encyclopedias of philosophy—all of which include reference to many exceedingly obscure *male* philosophers. If women philosophers are uniquely fitted to contribute to an understanding of women's experience, our women students have never been able to benefit from their contributions. Nor have women students been able to get a sense of themselves as philosophers or as capable of doing philosophy from them. Indeed, women philosophers' absence from courses has signalled to women students that women simply do not or cannot do philosophy.

So women students do not get women's contributions to philosophy out of their philosophy courses. But they *do* get *men's* contributions to philosophy out of their courses. Many of these men's contributions, however, leave much to be desired. For one thing, they contain very negative characterizations of women. Even Plato, one of the most egalitarian of philosophers—who holds that women and men possess the same virtues (though, generally speaking, women possess them to a lesser degree), and, as a consequence, should have the same opportunities for education and governmental and other positions in an ideal state (*Republic* V)—even Plato characterizes women in very negative terms. For example, he suggests that women, unlike men, are incapable of controlling their emotions. (Thus, he says: "When in our own lives some affliction comes to us you are aware that we plume ourselves . . . on our ability to remain calm and endure, in the belief that this is the conduct of a man, and [giving in to grief] that of a woman."[2]) Again, Plato suggests that women have more concern for their bodies than for their souls, whereas men have more concern for their souls. (For example, he says that the most proper penalty for a soldier who surrenders to save his body, when he should be willing to die out of the courage of his soul, is for him to be turned into a woman.[3]) What Plato seems to mean by these claims[4] is that *womanly* men and women—that is, people who behave in the ways we expect women to behave—are incapable of controlling their emotions and have more concern for their bodies than for their souls; whereas, *manly* men and women—that is, people who behave in the ways we expect *men* to behave—do not have these problems. In short, it is the womanly natures (souls) imprisoned in (some) male and (almost all) female bodies that are inferior, according to Plato, not simply all women.

---

[1] See, e.g., Linda Lopez McAlister's "Some Remarks on Exploring the History of Women in Philosophy," and Mary Ellen Waithe's "On Not Teaching the History of Philosophy," in *Hypatia's Special Issue: The History of Women in Philosophy*, IV, 1 (Spring, 1989).

Aristotle[5] is more negative. According to him, all women have the same inferior womanly nature—it is given by biology. Indeed, women are simply "misbegotten males," the result of failures in the "concocting" processes involved in reproduction. Had the failures not occurred, women would be men, genuine human beings; as it is, they are deformed men—deficient human beings—rather like eunuchs, except that the deficiency in this case lies in the rational part of their souls. Whereas in men the rational part of the soul naturally rules over the irrational part, in women the deliberative capacity of the rational part is easily overruled by the irrational part. As a consequence, women must be ruled by men.

Aquinas[6] is more negative still. Not only are women physiologically inferior to men, misbegotten males, but from the relative weakness of their reason follows their greater susceptibility to the disorder of sin. Women are thus men's moral, physical, and intellectual inferiors, and are thus always subject to them, with no possibility of escape. Alas, whereas the end and essence of man is intellectual activity, the sole reason woman was created was for purposes of reproduction, and it is only woman's reproduction and care of offspring that is of real assistance to man. In all other things man is better served by other men.

---

[2] *Republic* 605c–d; and cf 395d–e, *Phaedo* 60a and 112d, and *Apology* 356.

[3] *Laws* 944e; and cf. *Timaeus* 42b–c, 76e, 91a.

[4] See Elizabeth Spelman, "Hairy Cobblers and Philosopher-Queens," in *Inessential Woman* (Boston: Beacon, 1988).

[5] See *De Generatione Animalium* 728a 17, 766a 26, 768a 6 seq, 775a 15, 788a 5, and *Politics* I.

[6] See *Summa Theologica*, Part 1, Question 92.

Kant[7] is hardly less negative. For him, the entire character of woman is an embodiment of beauty, while man's nature is essentially noble or sublime. What is more, woman's understanding is a facile understanding as opposed to the deep understanding of man. Deep reflection and sustained contemplation, laborious study and rigorous disciplines like philosophy, are not suitable to her, and destroy the excellences peculiar to her sex. Virtue is best instilled in her when it is equated with beauty, vice with ugliness.

> Nothing of *ought*, nothing of *must*, nothing of *due*. All orders and all surly compulsion are to women insupportable. They do something but because they are pleased so to do, and the art consists but in making that which is good pleasing to them. I hardly believe that the fair sex are capable of principles. . . . Instead of which, however, Providence hath implanted in their breasts humane and benevolent sentiments, a fine feeling for becomingness, and a complaisant soul (*ibid.*).

And Nietzsche:[8] Woman is boring, stupid, deficient in taste, and insensitive to the truth. "Her great art is falsehood, her chief concern is appearance and beauty. . . ." She is "a possession . . . a being predestined for service and accomplishing her mission therein. . . ." Her first and last function is to bear robust children.

And the list of negative characterizations of women in men's contributions to philosophy goes on and on.

Note that these negative characterizations of women are each tied to a correspondingly positive characterization of men. Man is, for example, more concerned with the soul while woman is more concerned with the body, or man is directed by principle while woman is directed by sentiment, or man is intelligent while woman is stupid. Note also, that there is a strong air of unreality in these sets of descriptions. Surely there are individual differences among men and among women, occasioned by differences of upbringing and education, individual aptitudes and experiences, and so forth, but such individual differences are not allowed to mar the neat contrasts. In short, we are being given ways to think about the sexes which have highly questionable descriptive force but potentially powerful prescriptive force nonetheless.

What connections do these ways of thinking about women and men have to other views of their authors? If there are *close* connections, if, that is, the unacceptable views of women and men of Plato, Aristotle, and the rest are somehow inscribed in fundamental ways in their ethical or political or metaphysical or epistemological views, then this will suggest a very negative answer to our question of what women students can get out of their philosophy courses. For then it will turn out that major classics of philosophy very deeply denigrate women and valorize men even while they adequately capture neither. And this is by no means an implausible possibility. To be sure, even a casual perusal of the philosophical views in question discloses interesting similarities between them and the unfortunate views of the sexes. Consider, for example, Plato's view of justice expressed in the *Republic.* Just as justice in the state is the performance by each individual of the work for which that individual is best suited, Plato tells us, justice in the individual is the performance by each part of the individual's soul of the work for which that part is best suited, where this internal order of the soul necessarily results in right behavior. And injustice, whether in the state or in the individual, is simply the opposite state of internal discord and faction. Thus, in the perfectly just person the rational part of the soul, with its wisdom grounded on immediate knowledge of the good, rules over the bodily appetites, which form the greatest part of each person's soul; and it rules over that part of the soul with the other, spirited, part as its auxiliary. And correspondingly, in the perfectly just state these perfectly just persons, whose deliberations and wise counsel, remember, will be grounded on immediate knowledge of the good, rule over all other persons in the state with the aid of the executive persons, the auxiliaries.

---

[7] See *Observations on the Feeling of the Beautiful and Sublime*, sect. 3.

[8] See *Beyond Good and Evil*, 232, 234, 238, and 239.

But, as we have seen, control over bodily appetites characterizes a *manly* soul, not a womanly soul (whether in a man's or a woman's body). Remember, womanly persons (whether male or female) are incapable of controlling their emotions, are more concerned with the body, and the like. Thus, manly and not womanly persons are just, and justice is a manly trait. Of course, only a few aristocratic persons *are* manly, and not all the unmanly may be womanly. As Plato says:

> . . . The great mass of multifarious appetites and pleasures and pains will be found to occur chiefly in children and women and slaves, and, among free men so called, in the inferior multitude; whereas the simple and moderate desires which, with the aid of reason and right belief, are guided by reflection, you will find only in a few, and those with the best inborn dispositions and the best educated (*Republic* 431).

Thus, only a few aristocratic persons are manly because only such persons have the requisite superior inborn dispositions and superior education to enable them to obtain the immediate knowledge of the good that allows them to rule themselves and others with wisdom. And thus, knowledge of the good and wisdom are also manly characteristics, along with justice.

Notice how similar these patterns of thinking are to some of the ways in which many (most?) people currently think. Women are not really capable of rational thought. They are only intuitive. Moreover, they are too emotional and flighty, and too concerned with how they look and other trivial matters to have anything really intelligent and deep and insightful to offer. The uneducated, also, do not have anything worthwhile to offer, and certainly no *knowledge*. Knowledge is something obtained at universities, or reported in research papers or at conferences. Knowledge is not possessed by persons who have spent relatively little time in school, no matter how many practical life experiences may have enriched their understanding. Knowledge, above all, is possessed by professionals and the learned, scientists and judges and scholars, and the like, and they remain predominantly male.

But this is unfair, you will say. What is wrong with suggesting, as Plato does, that justice and knowledge and wisdom are characteristics of manly, not womanly, persons? It has absolutely no ill effects for women. After all, women can be manly too, as we have noted, and most men are *not* manly. So nothing is denied to women in affirming that justice and knowledge and wisdom are characteristics of manly persons. I have two replies. What is wrong with suggesting that justice and knowledge and wisdom are characteristics of manly, not womanly, persons is, first, that we thereby associate justice and knowledge and wisdom with men and the ways we expect men to behave, and dissociate them from women and our expectations of them; and, second, and relatedly, that we thereby suggest that the manly is all good and the womanly all bad. Let me clarify this second reply, taking justice as my example. Plato's concept of justice in the *Republic* does not seem to eschew what we might call traditional feminine virtues—concern for others, nurturance, cooperativeness, the willingness to make compromises, self-sacrifice. Indeed, the guardians, we are told (*Republic* 462–3), will look upon each other "as brother or sister, father or mother, son or daughter, grandchild or grandparent," and treat them all accordingly. And they "will have all their feelings of pleasure or pain in common": "In our community, then, above all others, when things go well or ill with any individual everyone will use that word 'mine' in the same sense and say that all is going well or ill with him and his." And "the people will look upon their rulers as preservers and protectors." So we end with a category system in which the manly seems to include all positive virtues and the womanly includes none, is simply a lack, an absence of the manly.

This is exactly the category system that occurs in Aristotle, though in a more damaging form. For now, the lack that is womanliness extends to all and only, not just most, women. It is, as I said before, a matter of biology. Women are simply "misbegotten males," the result of failures in the "concocting" processes involved in reproduction. Their reason is defective, stunted, and limited, in comparison to men's full reason. But the good life for human beings, Aristotle's subject in such books as the *Nicomachean Ethics* and the *Politics,* lies in the proper function and full development of that which distinguishes human beings from all other beings, that which makes them human. And this is their reason. As a result, women, along with those who most resemble them, children and slaves, are not capable of the human good that Aristotle describes. The *Nicomachean Ethics* and the *Politics* do not even apply to them.

Again, consider Kant on the subject of what gives actions moral worth.[9] According to Kant, the moral worth of an action resides in the motives out of which the action is performed. More particularly, to have moral worth an action must be done out of duty. Thus, for Kant, concern for others, sympathy, compassion, loving self-sacrifice, and the like are all irrelevant to the moral worth of an action and, ultimately, to the moral worth of a person's character, the highest worth a person can have. Indeed, it is only insofar as a person acts out of duty that a person can be described as moral at all. To act out of duty, moreover, necessarily involves acting out of principle, and it is one of the main aims of Kant's moral philosophy to identify a universal principle that can serve as the cornerstone for all morally right action.

But of course, according to Kant, this moral philosophy can only apply to men. For according to Kant, as we have seen, women are not capable of acting out of principle; they act out of "humane and benevolent sentiments" or out of a sense of "becomingness" or out of complaisance, but never out of duty or deep reflection. Thus, only men can have moral worth, the highest form of character there is. Kant's idea of moral worth, in fact, simply encapsulates his idea of the masculine, and excludes his idea of the feminine. And in a society like ours, in which women are still encouraged to be emotional and caring and concerned with others, and men are still encouraged to be unemotional and uninvolved and abstractly rational, Kant's evaluations of moral worth can be very damaging indeed.

---

[9] This example is taken from Jean Grimshaw, *Philosophy and Feminist Thinking* (Minneapolis: Minnesota UP, 1986)

I have been suggesting that the unacceptable ways of thinking about women and men that characterize many of the male philosophers of our philosophical tradition may have significant connections to their other philosophical views—may, in fact, be inscribed in fundamental ways in those other philosophical views. But what if I am wrong about this? What if there are no very significant or extensive or numerous connections between our male philosophers' unfortunate views of men and women and their other philosophical views? What if, in fact, their views of women and men have no conceptual connections at all to their other philosophical views? This may not be a happy result for our philosophers either, since a neutral stand toward the unacceptable views of men and women may be a condemnation in itself (compare: What if the ethical, or political, or epistemological, or metaphysical philosophies in question were neutral—as Plato's and Aristotle's are *not*—with regard to racism?) The only happy result for our philosophers, it seems, would be if the rest of their philosophies required a *negative* stand toward their unacceptable views of women and men. But alas, that possibility is the most implausible of all!

What is the upshot? We tell our students philosophy will teach them how to think and express themselves clearly, make them more critical of customary beliefs and assumptions, give them a general framework with which to help them make sense of their experiences, give them guidance in everyday life, and the like. Our women students, especially, *need* these skills. For the message is there in children's stories and textbooks (where women are the nurses, the mothers, the farmers' wives, but never the farmers, or where women are simply invisible), in advertising (where women are depicted as completely absorbed by the appearance of their nails or the appearance of their bathroom bowls), in the positions of greatest authority and status and power in society (all occupied by men), and even in our language (where *he* constitutes humanity, and *she* constitutes his little helpmate). Women and men are *still* raised in our society to see women as less capable and less valuable than men, intellectually, physically, and frequently even morally. And women must *deal* with that message of inferiority, and *combat* it, in themselves as well as in those who underpay them, fail to employ or promote them, sexually harass or rape them, and patronize and belittle them. But philosophy does not give women the wherewithal to do this. Indeed, far from pointing women toward a more adequate set of conceptions, philosophy—at least a goodly number of the classics of philosophy —*may* simply *further reinforce* the same message of women's inferiority. Nor does contemporary philosophy seem to do much better, since the issues it deals with so frequently have their roots in this same tradition, or have so little clear practical consequence (contemporary philosophy, for example, explores, too frequently, issues like whether I am free at time *t* to wiggle my finger at time *t*, or whether I am the same person as Smith even if I and two others were just constructed with Smith's old brain traces and exact copies of his body, or whether I know that Smith is at the library even if my belief is based on Jones' telling me what he thought was a falsehood, and . . . but you get the picture!)

Well, what can we do? I think if we are going to continue telling our students how enriching and valuable philosophy is, we are going to have to make some substantial changes in the philosophy curriculum. For one thing, I think we will have to include some of the recent feminist critiques of our philosophical tradition within our history of philosophy course reading matter, or at least devote class time to surveying this material. This will at least make our students aware of, and thereby somewhat immune to, but also analytically equipped to deal with, some of the misogyny and androcentrism operating within the philosophical tradition, though it will not give students any kind of alternatives to this tradition. To give them such alternatives, whether in terms of female role models or in terms of female-positive perspectives and valuings, I think we will have to integrate into our history courses some of the new scholarship now appearing regarding women philosophers from various periods in history. Notice, however, that including within our courses feminist critiques of our tradition as well as work by women philosophers will have to mean *excluding* some of the *men* philosophers who used to be included, or at least some of their work—there just will not be enough time to do it all! We will thus have to ask ourselves very seriously what our aim is in teaching the history of philosophy, and philosophy in general, what we hope to accomplish with our students, why the works we have covered in the past were thought to be truly valuable and enriching, and what is truly valuable and enriching about them, especially for contemporary students, female as well as male. We shall, of course, have to ask these questions of any possible new inclusions as well. We shall also have to consider the relations among our various thinkers—who influenced or further developed whom, who is needed to understand whom, who needs to be dropped or added because someone else has been dropped or added, and the like. Bringing in particular women figures, after all, may necessitate bringing in other women or men before them whom we never included in the curriculum before, just to prepare the way for the women. The result, I think, will be significant changes in the history of philosophy curriculum.

At this point, some of you might be feeling a little queasy. Leave out some of the most important names and/or some of the most important works in philosophy—for those are the only works we ever assigned, right?—for some heretofore unknowns, just to right the gender message and the gender balance in the curriculum? Is it really worth it? So let us try a little thought experiment: What if you were visiting for the term at a predominantly Black university? Would you include, say, *only* Plato and Aristotle on your syllabus? Imagine yourself, for example, standing day after day in front of fifty young, African-American women and men, the only white in the room, giving explanatory accounts of such topics as Plato's allegorical myth of the metals (the "noble lie," remember, that everyone is born either golden, silver, or iron and brass, and hence, needs to be promptly and permanently assigned to a corresponding station in society), or Aristotle's metaphysical principle that there is in the constitution of the universe a hierarchy in everything—there is always a ruler and a ruled. And imagine that such topics are all that you ever did. No critiques focused on racial issues, rather than the usual metaphysical and epistemological issues. No alternative perspectives. No African or African-American thinkers. Just Plato and Aristotle. Would you be at all uncomfortable, embarrassed, or ashamed teaching a philosophy of subjugation and nothing else to those who suffer from it? And if you would, then why do you now choose to teach only the likes of Plato and Aristotle to women? Perhaps you will say you can teach these works, and leave the really bad parts out. But that just gives students the thought patterns (or reinforces the thought patterns already in them) without showing them the logical consequences of those thought patterns. Is that really what you want?

But history of philosophy courses are not the only ones that could profit from change. Philosophy of science courses, for example, need to stop portraying scientists as disembodied minds, free of all human characteristics like gender, race, class, and political outlook, who experiment, explain, theorize, and validate in a perfectly detached, dispassionate, "objective" manner. They need, among other things, to inform students of the subtle and not so subtle ways in which women have been largely excluded from the scientific enterprise, and the epistemological and other consequences this has had. Philosophy of mind courses need to consider not only the question of what a person fundamentally is, but also, what men and women fundamentally are, to what extent and in what ways they are similar and different, and to what extent, as a result, women and men should fill the same roles in society, or to what extent the whole role structure should be changed. Metaphysics courses should consider, along with the abstract issues pertaining to the question of freedom and determinism, such concrete issues as the effects on self-development and action of gender socialization and interrelated race, class, and age stereotypes. Political philosophy courses need to give prominent attention to issues of gender and the family and their relations to the more traditional concerns covered in such courses. The proposals could go on and on.

Note that I am suggesting these curricular changes for *every* philosophy program: it is just no good to say, therefore, that *we* (in this or that department) should not make these changes because then we would not be adequately preparing our students for the other philosophy departments in which they might study. And note some of the other changes that would or might ensue. The students who become attracted to philosophy and then go on in philosophy might become more diversified, and philosophy might be enriched thereby. At the very least, recruiting women philosophers may become an easier task, since more women may become committed to philosophy. The ways in which we evaluate and fund philosophical research (e.g., in the history of philosophy) will have to change in order to support the new curricular requirements. And so, with that, the focus of doctoral research, hiring policies, and the like.

In everyday speech people distinguish between those issues and concerns which are of genuine interest and importance and use, and those issues which are "merely academic." With notable exceptions, philosophical issues and concerns have tended to be assigned an honored place within the latter category. One needs only to recall with what scorn and sarcasm newspaper journalists have reported the goings-on at American Philosophical Association annual meetings. For their part, philosophers, when they have agreed at all with this less-than-glowing assessment of philosophy, have tended to consider it the result of such things as philosophy's forever spawning new areas of science and becoming impoverished thereby. And, at any rate, philosophers have extolled the valuable critical contributions they have made and continue to make to their students and whoever else will listen. What I have been suggesting, however, is that philosophy may be less "merely academic" than people have thought, though relatedly, it may also be less critically valuable than philosophers have thought. I have also been suggesting that philosophy *can* be, ultimately, a powerful force for the good, not "merely academic" or impoverished at all. This is what I mean by "philosophy in a different voice," philosophy in a more inclusive, more responsive, *set* of voices.

JANET A. KOURANY

University of Notre Dame

FEMINISM AND THE CANON*

JANET A. Kourany questions the utility of the present philosophical curriculum for female students on the grounds of the sexist assumptions held by many philosophers who figure in our tradition (her examples include Plato, Aristotle, and Kant) *and* on the grounds of the exclusion of women's contributions to the tradition. My response will address both sets of concerns. I inquire as to the ways in which certain philosophers' views have been distorted by sexism; I do this to determine the kind of "connections" (Kourany's term) which may exist between the bias in question and the core doctrines of, for example, Plato and Kant. My own conclusions differ from Kourany's insofar as I want to pursue the issue of the *centrality* of the bias to their philosophies; it may be that this prejudice is itself lodged within other theoretical assumptions. Even if Kant defines women and moral excellence in such a way that never the twain shall meet (as Kourany argues), can it not be the case that, correcting for his sexism, Kantian ethics may yet have its merits, for women and men? Then again, it may be that Kant's infamous separation of inclination and reason is vulnerable to objection by women and men alike. Just so, Plato's identification of virtue, in the *Republic,* with the so-called "manly soul," and of moral weakness with the "womanly soul," is certainly objectionable. But Socrates's position on the role of reason in relation to the body, *on* which these distinctions rest, is itself open to serious question. I mean to indicate that it is not just sexist bias that is the problem here, in Plato or Kant: it is the philosophical doctrines in which these biases are imbedded. Thus, feminist criticism leads us to central philosophical questions within these thinkers.

This raises the vexed question of feminist philosophy's relation to the accepted tradition. I want to argue that much in the tradition can serve well male and female alike; likewise, even the tradition's unquestionable misogyny, and that misogyny's varied theoretical underpinnings, can be instructive—once these are thematized. Thus, one place for feminist questioning is in the classroom, where the dialectic of, for example, a Platonic dialogue, can be extended to issues not perhaps anticipated by Plato himself. The curricular reform that Kourany urges upon us must needs proceed with great

---

* Abstract of a paper to be presented in an APA symposium on Philosophy in a Different Voice, December 29, commenting on a paper by Janet A. Kourany. See this JOURNAL, this issue, 557–67, for her contribution.

0022-362X/91/8810/568–9

caution, I argue. However much the present canon of philosophical texts embodies certain biases, it is still true that what philosophy most "promises" its students, to use Kourany's phrase, is understanding and skill in thinking, and the concepts with which feminist thought must grapple originate within the classical texts of the tradition. To lack knowledge of these might well impair further feminist inquiry and self-understanding. Insofar as feminist thought goads us to question and criticize our authorities, to be persuaded by reason and not by force or mere rhetoric, it is in deep sympathy with the best of what has been called philosophy in the Western tradition. And, on the grounds of truth, if feminist thought can contest or even refute what is *not* the best in the tradition, so much the better. It may be one of the distinct contributions of feminist thought to philosophy that it teaches us about what we have assumed heretofore, that it uncovers what has been buried, for bad or good, within accepted wisdom. If it shows as folly what has long been deemed truth within the canon, that may be an irony even Socrates could love.

PAMELA HALL

Emory University

JOHNSONIAN PRIZE IN PHILOSOPHY

The editors are pleased to announce that the winner of the Johnsonian Prize in Philosophy for 1990 is Christine Swanton, Senior Lecturer of Philosophy, University of Auckland, for her manuscript, *Freedom: A Coherence Theory,* which will be published soon by Hackett Publishing, the co-sponsor of the prize.

Previous Johnsonian prize winners have been: 1980, Daniel Bonevac, *Reduction in the Abstract Sciences* (1981); 1981, Igal Kvart, *A Theory of Counterfactuals* (1986); 1982, Moreland Perkins, *Sensing the World* (1983); 1984, Robert J. Dunn, *The Possibility of Weakness of Will* (1986); 1985, Harry van der Linden, *Kantian Ethics and Socialism* (1988); 1986, Clyde L. Hardin, *Color for Philosophers* (1988); 1988, Vann McGee, *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth* (1991). Also, honorable mention in 1981 went to Charles B. Guignon, *Heidegger and the Problem of Knowledge* (1983); and in 1986 to Dionysios Anapolitanos, *Leibniz: Representation, Continuity, and the Spatiotemporal.* All of these titles will be on display with Hackett at the APA Division Meetings in 1991–92.

Announcement of the ninth and final prize winner will be made soon in an issue of the JOURNAL.

PLEASE NOTE: The JOURNAL and Hackett Publishing will not be offering the prize after this year, 1991.

AMERICAN PHILOSOPHICAL ASSOCIATION
EASTERN DIVISION
EIGHTY-EIGHTH ANNUAL MEETING

MARRIOTT MARQUIS HOTEL, NEW YORK, NEW YORK
DECEMBER 27–30, 1991

PROGRAM

FRIDAY, DECEMBER 27

1:00 P.M.–10:00 P.M. REGISTRATION, *Fifth Floor*
1:00 P.M. EXECUTIVE COMMITTEE MEETING, *Barrymore*

PLACEMENT INFORMATION

For interviewers, 1:00 P.M.–7:00 P.M., *Duffy/ Columbia*
For candidates, 1:00 P.M.–7:00 P.M., *Gramercy/Herald/Soho*

EVENING (8:00–11:00 P.M.)

Group Meetings (See Appendix for Programs)

8:00 LENSGRINDER'S SOCIETY, *Odets/Wilder*
8:00 SOCIETY FOR THE STUDY OF PROCESS PHILOSOPHIES *Kern/Sullivan*
8:00 SOCIETY FOR ASIAN AND COMPARATIVE PHILOSOPHY *Juliard/Imperial*
8:00 PERSONALIST DISCUSSION GROUP, *Broadhurst/Belasco*
8:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Schubert/Uris*
8:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Carnegie/Lyceum*

SATURDAY, DECEMBER 28

9:00 A.M.–7:00 P.M. REGISTRATION, *Fifth Floor*
9:00 A.M.–5:00 P.M. BOOK DISPLAYS, *Westside Ballroom*
9:00 A.M.–5:00 P.M. *JOURNAL OF PHILOSOPHY* DISPLAY, *Fifth Floor*

CHILD CARE arrangements. See the October issue of the *APA Proceedings* for further details.

PLACEMENT INFORMATION

For interviewers, 8:30 A.M.–7:00 P.M., *Duffy/Columbia*
For candidates, 8:30 A.M.–7:00 P.M., *Gramercy/Herald/Soho*

Saturday, December 28, Morning

MORNING (9:00–11:00)

Special Panel Arranged by the APA Committees on the Status of Women and on Philosophy and Law

Boothe/Edison/Alvin

Topic: Pregnancy in the Hazardous Workplace: In the Wake of *Johnson Controls*

I. Uma Narayan, "The Gendered Nature of Industrial Health Hazard Policies: Implications for Third-world Women Workers"

II. John Robertson, "Prenatal Maternal Duties to Offspring in and out of the Workplace"

III. Joan C. Callaghan, "Hazardous Workplaces and Prenatal Harm: Let's Get the Lead out"

IV. Lawrence D. Houlgate, "Familial Rights and Responsibilities: A Balancing Act If There Ever Was One"

Group Meetings (See Appendix for Programs)

9:00 SOCIÉTÉ AMÉRICAINE DE PHILOSOPHIE DE LANGUE FRANÇAISE *O'Neill*
9:00 RADICAL PHILOSOPHY ASSOCIATION, *Odets/Wilder*
9:00 AMERICAN SOCIETY FOR VALUE INQUIRY, *Miller*
9:00 GANDHI-KING SOCIETY, *Albee*
9:00 WORLD PHENOMENOLOGY INSTITUTE, *Hart*
9:00 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, *Brecht*
9:00 SOCIETY FOR PHILOSOPHY AND TECHNOLOGY, *Plymouth/Royale*
9:00 SOCIETY FOR PHILOSOPHIC STUDY OF THE CONTEMPORARY VISUAL ARTS, *Carnegie/Lyceum*

LATE MORNING (11:15–1:15)

11:15 SOCIETY FOR INDIAN PHILOSOPHY AND RELIGION, *Gilbert*
11:15 ASSOCIATION FOR PHILOSOPHY AND LIBERATION, *O'Neill*
11:15 SOCIETY FOR THE ADVANCEMENT OF AMERICAN PHILOSOPHY *Odets/Wilder*
11:15 SØREN KIERKEGAARD SOCIETY, *Miller*
11:15 SARTRE CIRCLE, *Albee*
11:15 SOCIETY OF PHILOSOPHERS AT WORK IN THE WORLD, *Hart*
11:15 SOCIETY FOR THE PHILOSOPHICAL STUDY OF MARXISM, *Brecht*

11:15 SOCIETY FOR WOMEN IN PHILOSOPHY, *Schubert/Uris*
11:15 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Kern/Sullivan*

AFTERNOON

2:00–5:00 Session I

A. Symposium: Logical Positivism: A Retrospective *Broadway Ballroom/North*
Chair: Warren Goldfarb
Speaker: Michael Friedman
Commentators: Richard Creath and David Weissman

B. Symposium: Post-Kantian Idealism, *Broadway Ballroom/South*
Chair: Tom Rockmore
Speakers: J. Daniel Breazeale and Robert Pippin
Commentator: Chernor Jalloh

C. Book Discussion: Kendall Walton's *Mimesis as Make Believe* *Salon B/C*
Chair: Arthur C. Danto
Speakers: Alan Goldman and Nicholas Wolterstorff
Respondent: Kendall Walton

D. Special APA Panel: Philosophical Perspectives on the Transformation of Eastern Europe and the Baltic States *Carnegie/Lyceum*
Chair: Ernest Sosa
Speakers: Erazim Kohak, Ignas Skrupskelis, and Mark Tushnet

E. Colloquium: Philosophy of Law, *Plymouth/Royale*
1. *Chair:* Andrew Altman
   Frederick Schauer, "Defeasibility Defeated"
   Commentator: Raymond Belliotti
2. *Chair:* Robert Paul Wolff
   Luc Bovens, "Contextual Pluralism and the Libertarian Paradox"
   Commentator: George Sher
3. *Chair:* Margaret Gilbert
   Elizabeth Kiss, "Is It Dangerous to Demystify Human Rights?"
   Commentator: Joseph Grassi

Saturday, December 28, Afternoon, Session I

F. Colloquium: Descartes, Leibniz, and the Divine *Palace/Winter Garden*

1. *Chair:* James A. Montmarquet
   David Ring, "Idea as the Form of a Thought in Descartes"
   Commentator: Paul Hoffman
2. *Chair:* Jan Cover
   Michael Murray, "Leibniz on Divine Foreknowledge and Human Freedom"
   Commentator: Graeme Hunter
3. *Chair:* William Mann
   Jeff Jordan, "The Problem of Divine Exclusivity"
   Commentator: Philip Quinn

G. Colloquium: Aristotle, *Music Box/Majestic*

1. *Chair:* John Russon
   Sibyl Schwarzenbach, "A Political Reading of the Reproductive Soul in Aristotle"
   Commentator: Sheldon Cohen
2. *Chair:* Allan Gotthelf
   James Celarier, "Reflections on Matter and Individuation in Aristotle"
   Commentator: Charlotte Witt
3. *Chair:* Michael Roth
   Robin Smith, "Aristotle on the Uses of Dialectic"
   Commentator: Roderick Long

H. Colloquium: Philosophy of Mind, *Kern/Sullivan*

1. *Chair:* Thomas Cook
   Irwin Goldstein, "A Critique of Analytic Functionalism"
   Commentator: Robert Burton
2. *Chair:* Willem de Vries
   Todd May, "The Limits of the Mental and the Limits of Philosophy: From Burge to Foucault and beyond"
   Commentator: Thomas Flynn
3. *Chair:* Lynne Baker
   Steven Mandelker, "An Argument against the Externalist Account of Intentional Content"
   Commentator: William G. Lycan

I. Special Symposium Arranged by the APA Committee on Blacks in Philosophy, *Broadhurst/Belasco*

Topic: "Critical Race Theory" in Jurisprudence and Philosophy
Chair: Anita Allen
Speakers: David Goldberg, Leonard Harris, Gerald Torres, and Patricia Williams

J. Special Session Arranged by the APA Committees on the Status of Women and on International Cooperation, *Boothe/Edison/Alvin*

Topic: Philosophy and Women in Developing Countries Today
Chair: Michele Moody-Adams

I. Xiaorong Li, "A Philosophical Reflection on Women's Status in China"
II. Laura Mues de Schrenk, "Gender, Class, and Social Development in Mexico"
III. Roop Rekha Verma, "Feminity, Equality, and Personhood"
IV. Nkiru Nzegwu, "Postcolonial Alienation: A Case for Traditional Women's Organization in Development"

Commentators: David Crocker and Martha Nussbaum

Group Meetings (See Appendix for Programs)

2:00 SOCIETY FOR BUSINESS ETHICS, *O'Neill*
2:00 INTERNATIONAL SOCIETY FOR CHINESE PHILOSOPHY, *Miller*
2:00 INTERNATIONAL SOCIETY FOR ENVIRONMENTAL ETHICS, *Albee*
2:00 SOCIETY FOR IBERIAN AND LATIN AMERICAN THOUGHT, *Hart*

LATE AFTERNOON (5:15–7:15)

5:15 CHARLES S. PEIRCE SOCIETY, *O'Neill*
5:15 SOCIETY FOR ANCIENT GREEK PHILOSOPHY, *Odets/Wilder*
5:15 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, *Albee*
5:15 SOCIETY FOR MACHINES AND MENTALITY, *Kern/Sullivan*
5:15 SOCIETY FOR PHILOSOPHY AND PUBLIC AFFAIRS, *Juliard/Imperial*
5:15 SOCIETY FOR THE STUDY OF HUSSERL'S PHILOSOPHY, *Broadhurst/Belasco*
5:15 JASPERS SOCIETY OF NORTH AMERICA, *Plymouth/Royale*
5:15 ASSOCIATION FOR THE PHILOSOPHY OF THE UNCONSCIOUS *Palace/Winter Garden*
5:15 ASSOCIATION OF PHILOSOPHY JOURNAL EDITORS, *Miller*

EVENING (7:30–10:30)

7:30 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING, *O'Neill*
7:30 SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, *Odets/Wilder*
7:30 INTERNATIONAL BERKELEY SOCIETY, *Miller*
7:30 INTERNATIONAL SOCIETY FOR UNIVERSALISM, *Albee*
7:30 NORTH AMERICAN KANT SOCIETY, *Hart*
7:30 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY, *Kern/Sullivan*
7:30 NORTH AMERICAN SOCIETY FOR SOCIAL PHILOSOPHY, *Brecht*

Saturday, December 28, Evening

7:30 AMERICAN SOCIETY FOR VALUE INQUIRY AND THE INTERNATIONAL SOCIETY FOR ENVIRONMENTAL ETHICS, *Juliard/Imperial*
7:30 ASSOCIATION FOR PHILOSOPHY OF EDUCATION, *Broadhurst/Belasco*
7:30 SOCIETY FOR WOMEN IN PHILOSOPHY, *Carnegie/Lyceum*
7:30 SOCIETY FOR ASIAN AND COMPARATIVE PHILOSOPHY, *Salon A*
7:30 LEIBNIZ SOCIETY OF NORTH AMERICA, *Salon B/C*
7:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Cantor/Jolson*
7:30 CONCERNED PHILOSOPHERS FOR PEACE, *Gilbert*
9:00 P.M.–12:00 Midnight RECEPTION, *Broadway Ballroom*

SUNDAY, DECEMBER 29

9:00 A.M.–7:00 P.M. REGISTRATION, *Fifth Floor*
9:00 A.M.–5:00 P.M. BOOK DISPLAYS, *Westside Ballroom*
9:00 A.M.–5:00 P.M. *JOURNAL OF PHILOSOPHY* DISPLAY, *Fifth Floor*

PLACEMENT INFORMATION

For interviewers, 8:00 A.M.–7:00 P.M., *Duffy/Columbia*
For candidates, 8:30 A.M.–7:00 P.M., *Gramercy/Herald/Soho*

MORNING

9:00–11:00 Session II

A. Symposium: The Philosophy of Music, *Salon B/C*
Chair: Kendall Walton
Speaker: Peter Kivy
Commentator: Philip Alperson

B. Symposium: Philosophy in a Different Voice *Broadway Ballroom/South*
Chair: Merold Westphal
Speaker: Janet Kourany
Commentator: Pamela Hall

C. Invited Paper, *Schubert/Uris*
Chair: James Ross
David Burrell, "Creation and 'Actualism': The Dialectical Dimension of Philosophical Theology"
Commentator: Robert M. Adams

D. Invited Paper, *Carnegie/Lyceum*
Chair: Drew Hyland
Charles Sherover, "The Conditions of Freedom"
Commentator: Edith Wyschograd

E. Colloquium: Rorty, *Plymouth/Royale*
  1. *Chair:* Michael Kelly
    Andrew Cutrofello, "Richard Rorty and the 'Foucauldian Left' "
    Commentator: Felmon Davis
  2. *Chair:* James Edwards
    Steven Hendley, "Putting Ourselves up for Question: A 'Postmodern' Critique of Rorty's 'Postmodernist Bourgeois Liberalism' "
    Commentator: Jay M. Van Hook

F. Colloquium: Metaphysics, *Palace/Winter Garden*
  1. *Chair:* Justin Schwartz
    Mark Rollins, "Picture Parsing"
    Commentator: John Brough
  2. *Chair:* John Post
    Douglas Ehring, "Property Persistence: What Universals Cannot Do but Tropes Can"
    Commentator: Donald Baxter

G. Colloquium: Care and Viability, *Music Box/Majestic*
  1. *Chair:* Willem Landman
    Patricia Ward Scaltsas, "Singular Moral Statements and the Ethic of Care"
    Commentator: Patricia James
  2. *Chair:* Jane Mary Trau
    Heather J. Gert, "Viability"
    Commentator: Meredith Michaels

H. Special Panel Arranged by the APA Committee on Precollege Instruction in Philosophy, *Kern/Sullivan*
  Topic: International Developments in Elementary School Philosophy
  Chair: Rosalind Ekman Ladd
  Panel Discussion with Matthew Lipman

I. Special Panel Arranged by the APA Committee on Teaching Philosophy in Two-year Colleges, *Miller*
  Topic: Teaching Ethics: What Should We Be Trying to Accomplish?
  Panel: Julien Farland, Eric Lindermayer (Moderator), and Neil Rossman

Sunday, December 29, Morning, Session II

J. Special Session Arranged by the APA Committee on the Status of Women, *Boothe/Edison/Alvin*

I. Helen Longino, Report from the Committee on the Status of Women

II. *Topic:* Sandra Lee Bartky's *Feminity and Domination*
Chair: Marilyn Friedman
Speakers: Bat-Ami Bar-On, Douglas Kellner, Kathryn Morgan, and Patrocinio Schweickart
Respondent: Sandra Lee Bartky

Group Meetings (See Appendix for Programs)

9:00 SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, *O'Neill*
9:00 HUME SOCIETY, *Odets/Wilder*
9:00 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY, *Albee*
9:00 WORLD PHENOMENOLOGY INSTITUTE, *Brecht*
9:00 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY *Juliard/Imperial*
9:00 SOCIETY FOR THE PHILOSOPHIC STUDY OF THE CONTEMPORARY VISUAL ARTS, *Hart*

11:30 A.M.–1:30 P.M. BUSINESS MEETING *Broadway Ballroom/North*

AFTERNOON

2:15–4:15 Session III

A. Symposium: The Epistemology of Modality, *Salon B/C*
Chair: Ruth Marcus
Speaker: Phillip Bricker
Commentator: Joseph Almog

B. Symposium: Aquinas on the Veracity of the Intellect *Broadway Ballroom/South*
Chair: Gareth Matthews
Speaker: John Jenkins
Commentator: Eleonore Stump

C. Invited Paper, *Schubert/Uris*
Chair: Marx Wartofsky
Berel Lang, "The Style of Method: Repressed Questions in the Genealogy of Philosophy"
Commentator: Gary Shapiro

D. Invited Paper, *Carnegie/Lyceum*
Chair: Richard Miller
Joseph Walsh, "The Impossible, but Necessary, Ethics of Revolution"
Commentator: Kevin Brien

E. Colloquium: Philosophy of Law, *Plymouth/Royale*
1. *Chair:* Lois Pineau
Heta and Matti Hayry, "Feinberg, Obscenity, and the Harmfulness of Artistic Pornography"
Commentator: Betsy Postow
2. *Chair:* Thomas Magnell
Frederick R. Kellogg, "Pragmatism and Liberalism: Two Distinct Theories of Law and Justice"
Commentator: Lucius Outlaw

F. Colloquium: Philosophy of Mind, *Palace/Winter Garden*
1. *Chair:* John Heil
Robert Van Gulick, "Understanding the Phenomenal Mind"
Commentator: Joseph Levine
2. *Chair:* John-Christian Smith
Beth Preston, "Behaviorism/Mentalism: The Only Game in Town?"
Commentator: Scott Davison

G. Colloquium: Heidegger, *Music Box/Majestic*
1. *Chair:* Calvin Schrag
Michael David Sickels, "The Intentionality of Being"
Commentator: John Drummond
2. *Chair:* Joseph Margolis
James Robert Quick, "Heidegger on Science: From Phenomenology to National Socialism and beyond"
Commentator: Jeremy Barris

H. Special Panel Arranged by the APA Committees on the Teaching of Philosophy, on the Status of Women, and on Blacks in Philosophy, *Odets/Wilder*
Topic: Transforming the Canon
Panel: David Crocker, Sandra Harding, Leonard Harris, and Eileen O'Neill

Sunday, December 29, Afternoon

Group Meetings (See Appendix for Program)

2:15 AYN RAND SOCIETY, *O'Neill*
2:15 SOCIETY FOR THE PHILOSOPHICAL STUDY OF MARXISM, *Miller*
2:15 SOCIETY FOR THE PHILOSOPHY OF CREATIVITY, *Albee*
2:15 RADICAL PHILOSOPHY ASSOCIATION, *Hart*
2:15 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Juliard/Imperial*
2:15 NORTH AMERICAN NIETZSCHE SOCIETY, *Kern/Sullivan*

4:30 PRESIDENTIAL ADDRESS, *Broadway Ballroom/North*
Introduction by Judith J. Thomson
Harry Frankfurt, "The Faintest Passion"

EVENING (6:30–9:30)

Group Meetings (See Appendix for Programs)

6:30 SOCIETY FOR REALIST/ANTIREALIST DISCUSSION, *Gilbert*
6:30 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING *O'Neill*
6:30 INTERNATIONAL SOCIETY FOR ENVIRONMENTAL ETHICS, SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, AND RADICAL PHILOSOPHY ASSOCIATION *Odets/Wilder*
6:30 SOCIETY FOR LESBIAN AND GAY PHILOSOPHY, *Miller*
6:30 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, *Albee*
6:30 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, *Hart*
6:30 SOCIETY FOR THE PHILOSOPHY OF SEX AND LOVE *Kern/Sullivan*
6:30 JASPERS SOCIETY OF NORTH AMERICA, *Juliard/Imperial*
6:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES *Broadhurst/Belasco*
9:00 P.M.–12:00 Midnight RECEPTION, *Broadway Ballroom*

MONDAY, DECEMBER 30

9:00 A.M.–12:00 Noon REGISTRATION, *Fifth Floor*
9:00 A.M.–2:00 P.M. BOOK DISPLAYS, *Westside Ballroom*
9:00 A.M.–12:00 Noon *JOURNAL OF PHILOSOPHY* DISPLAY *Fifth Floor*

PLACEMENT INFORMATION

For interviewers, 8:30 A.M.–12:00 Noon, *Duffy/Columbia*
For candidates, 9:00 A.M.–12:00 Noon *Gramercy/Herald/Soho*

MORNING

9:00–11:00 Session IV

A. Symposium: Feminist Critiques of Equality
Broadway Ballroom/North
Chair: Robin Schott
Speaker: Martha Minow
Commentator: Eva Kittay

B. Symposium: Can Biological Teleology Be Naturalized?
Broadway Ballroom/South
Chair: Steven Wagner
Speaker: Mark Bedau
Commentator: Mohan Matthen

C. Invited Paper, *Salon B/C*
Chair: Bruce Aune
Susan Haack, "Causical Foundherentism"
Commentator: Peter Hare

D. Invited Paper, *Carnegie/Lyceum*
Chair: Dennis Schmidt
Kathleen Wright, "Hegel and Heidegger on Antigone or 'Killing Her Softly' "
Commentator: Martin Donougho

E. Invited Paper, *Broadhurst/Belasco*
Chair: Robert Corrington
Beth Singer, "Buchler and the Question of Foundations"
Commentator: Kathleen Wallace

F. Colloquium: Derrida and Narratives, *Plymouth/Royale*
1. *Chair:* Gayle Ormiston
Barry Allen, "The Secondary Sign: Derrida's Semiology"
Commentator: J. Claude Evans
2. *Chair:* David Pellauer
Margaret Urban Walker, "Responsibility, Narratives, and Integrity"
Commentator: Andrew Piker

G. Colloquium: Socrates and Plato, *Palace/Winter Garden*
1. *Chair:* Scott Calef
Naomi Reshotko, "The Socratic Theory of Motivation"
Commentator: Richard Parry
2. *Chair:* Thomas Brickhouse
Theodore Scaltsas, "A Semantic Impossibility in the Third-man Argument"
Commentator: Sherry Blum

Monday, December 30, Morning, Session IV

H. Colloquium: Promises and Public Policy, *Music Box/Majestic*
  1. *Chair:* Samuel Freeman
    Dennis M. Patterson, "Scanlon on Promises and Practices: A Tale of Incommensurables"
    Commentator: Thomas Kearns
  2. *Chair:* Robert Ackermann
    Andrew Buchwalter, "Philosophy, Culture, and Public Policy"
    Commentator: Adrian Piper

I. 9:00–12:00 Special Panel Arranged by the APA Committee on International Cooperation, *Boothe/Edison/Alvin*
  Topic: Philosophy in China Today
  Chair: Robert Cummings Neville
    I. Roger Ames, "Memories of Angus Graham"
    II. Weihang Chen, "Philosophical Scholarship in China Today"
    III. Chung-ying Cheng, "Rational Critique and Self-understanding in Philosophy in China Today"
    IV. Vincent Shen, "Creativity as Synthesis of Contrasting Wisdoms: Main Currents, Evaluation, and Future Prospect of Contemporary Chinese Philosophy"
    V. Lik Kuen Tong, "Contemporary Neo-Confucianism and Its Critics"

9:00 EXECUTIVE COMMITTEE MEETING, *Barrymore*

Group Meetings (See Appendix for Programs)

9:00 ASSOCIATION FOR PHILOSOPHY OF EDUCATION, *Odets/Wilder*
9:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Kern/Sullivan*
9:00 INTERNATIONAL SOCIETY FOR NEOPLATONIC STUDIES, *Hart*
9:00 WORLD PHENOMENOLOGY INSTITUTE, *Brecht*

LATE MORNING

11:15 RADICAL PHILOSOPHY ASSOCIATION, *Miller*
11:15 SARTRE CIRCLE, *Albee*
11:15 SOCIETY FOR WOMEN IN PHILOSOPHY, *Carnegie/Lyceum*
11:15 SOCIETY FOR REALIST/ANTIREALIST DISCUSSION, *O'Neill*
11:15 A.M.–12:15 P.M. BUSINESS MEETING (Second Session, if needed), *Broadway Ballroom/North*

AFTERNOON

1:30–4:30 Session V

A. Symposium: Knowledge, Power, and Science
Broadway Ballroom/North
Chair: Patrick Heelan
Speakers: Joseph Rouse and Helen Longino
Commentator: Philip Kitcher

B. Symposium: What Makes Life Worth Living?
Broadway Ballroom/South
Chair: Laurence Thomas
Speakers: John McDermott and Karsten Harries
Commentator: Jay Rosenberg

C. Book Discussion: Alan Gibbard's *Apt Feelings, Wise Choices*
Salon B/C
Chair: Charles Scott
Speakers: Amelie Rorty and Simon Blackburn
Respondent: Alan Gibbard

D. Colloquium: Locke, Hume, and Kant, *Plymouth/Royale*
1. *Chair:* John Robison
   John D. Greenwood, "Personal Identity, Locke's Intuition, and Psychological Atomism"
   Commentator: Vere Chappell
2. *Chair:* Anne Jacobson
   Andrew Ward, "Fictions and Hume's Idea of Necessary Connection"
   Commentator: Georges Dicker
3. *Chair:* Donald Becker
   James O'Shea, "Hume and Kant on Duration and Identity"
   Commentator: Michael Hodges

E. Colloquium: Protocols, Logical Consequences, and Pronouns
Palace/Winter Garden
1. *Chair:* Thomas Oberdan
   Thomas E. Uebel, "The Logic of Neurath's Protocols"
   Commentator: Paul Moser
2. *Chair:* G. Lee Bowie
   Gila Sher, "Did Tarski Commit 'Tarski's Fallacy'?"
   Commentator: Patricia Blanchette
3. *Chair:* Mark Crimmins
   Thomas J. McKay, "Unbound Pronouns"
   Commentator: James Higginbotham

Monday, December 30, Afternoon, Session V

F. Colloquium: James and Dewey, *Music Box/Majestic*

1. *Chair:* Charlene Seigfried
   Nobuo Kazashi, "James on the 'Horizon': Where Pragmatism and Phenomenology Meet: A Reappraisal through a Comparison with Merleau-Ponty" (William James Prize Essay)
   Commentator: Bruce Wilshire
2. *Chair:* James Campbell
   Jennifer Welchman, "Is Dramatic Rehearsal Dewey's Theory of Moral Deliberation?"
   Commentator: Victor Kestenbaum
3. *Chair:* Thomas Carlson
   Hans Seigfried, "Dewey's Experiments and the Business of Philosophers"
   Commentator: Russell Goodman

G. Colloquium: Knowledge, Time, and Deliberation *Juliard/Imperial*

1. *Chair:* Richard Feldman
   Crispin Sartwell, "Why Knowledge Is Merely True Belief"
   Commentator: Jonathan Vogel
2. *Chair:* James Fetzer
   John Hawthorn, "Making Waves"
   Commentator: Ned Markosian
3. *Chair:* Andrew Cling
   David P. Hunt, "Does Deliberation Presuppose Ignorance?"
   Commentator: Carl Ginet

PROGRAM PARTICIPANTS
EIGHTY-EIGHTH ANNUAL MEETING

ACKERMANN, Robert, University of Massachusetts/Amherst	IV-H
ADAMS, Robert M., University of California/Los Angeles	II-C
ALLEN, Anita, Georgetown University	I-I
ALLEN, Barry, McMaster University, Canada	IV-F
ALMOG, Joseph, University of California/Los Angeles	III-A
ALPERSON, Philip, University of Louisville	II-A
ALTMAN, Andrew, George Washington University	I-E
AMES, Roger, University of Hawaii/Manoa	IV-I
AUNE, Bruce, University of Massachusetts/Amherst	IV-C
BAKER, Lynne, Middlebury College, Vermont and University of Massachusetts/Amherst	I-H
BAR-ON, Bat-Ami, State University of New York/Binghamton	II-J
BARRIS, Jeremy, Marshall University	III-G
BARTKY, Sandra Lee, University of Illinois/Chicago	II-J
BAXTER, Donald, University of Connecticut	II-F
BECKER, Donald, University of Texas/Austin	V-D
BEDAU, Mark, Reed College, Oregon	IV-B
BELLIOTTI, Raymond, State University of New York/Fredonia	I-E
BLACKBURN, Simon, University of North Carolina/Chapel Hill	V-C
BLANCHETTE, Patricia, Yale University	V-E
BLUM, Sherry, University of Texas/Austin	IV-G
BOVENS, Luc, University of Colorado/Boulder	I-E
BOWIE, G. Lee, Mount Holyoke College, Massachusetts	V-E
BREAZEALE, J. Daniel, University of Kentucky/Lexington	I-B
BRICKER, Phillip, University of Massachusetts/Amherst	III-A
BRICKHOUSE, Thomas, Lynchburg College, Virginia	IV-G
BRIEN, Kevin, Washington College, Massachusetts	III-D
BROUGH, John, Georgetown University	II-F
BUCHWALTER, Andrew, University of North Florida	IV-H
BURRELL, David, University of Notre Dame	II-C
BURTON, Robert, University of Georgia	I-H
CALEF, Scott, Xavier University of Louisiana	IV-G
CALLAGHAN, Joan C.	p 572
CAMPBELL, James, University of Toledo	V-F
CARLSON, Thomas, Howard University	V-F
CELARIER, James, University of Maryland	I-G
CHAPPELL, Vere, University of Massachusetts/Amherst	V-D
CHEN, Weihang, Wuhan University, China	IV-I
CHENG, Chung-ying, University of Hawaii/Manoa	IV-I
CLING, Andrew, University of Alabama/Huntsville	V-G
COHEN, Sheldon, University of Tennessee/Knoxville	I-G
COOK, Thomas, Rollins College, Florida	I-H
CORRINGTON, Robert, Drew University	IV-E
COVER, Jan, Purdue University	I-F

CREATH, Richard, Arizona State University	I-A
CRIMMINS, Mark, Cornell University	V-E
CROCKER, David, Colorado State University	I-J, III-H
CUTROFELLO, Andrew, Saint Mary's College, Indiana	II-E
DANTO, Arthur C., Columbia University	I-C
DAVIS, Felmon, State University of New York/Albany	II-E
DAVISON, Scott, University of Notre Dame	III-F
DE SCHRENK, Laura Mues, University of Mexico	I-J
DE VRIES, Willem, University of New Hampshire	I-H
DICKER, Georges, State University of New York/Brockport	V-D
DONOUGHO, Martin, University of South Carolina	IV-D
DRUMMOND, John, Mount Saint Mary's College, Maryland	III-G
EDWARDS, James, Furman University	II-E
EHRING, Douglas, Southern Methodist University	II-F
EVANS, J. Claude, Washington University	IV-F
FARLAND, Julien, Middlesex Community College, Massachusetts	II-I
FELDMAN, Richard, University of Rochester	V-G
FETZER, James, University of Minnesota/Duluth	V-G
FLYNN, Thomas, Emory University	I-H
FRANKFURT, Harry, Yale University	Presidential Address
FREEMAN, Samuel, University of Pennsylvania	IV-H
FRIEDMAN, Marilyn, Washington University	II-J
FRIEDMAN, Michael, University of Illinois/Chicago	I-A
GERT, Heather J., Texas A & M University	II-G
GIBBARD, Alan, University of Michigan	V-C
GILBERT, Margaret, University of Connecticut	I-E
GINET, Carl, Cornell University	V-G
GOLDBERG, David, Arizona State University	I-I
GOLDFARB, Warren, Harvard University	I-A
GOLDMAN, Alan, University of Miami	I-C
GOLDSTEIN, Irwin, Davidson College, North Carolina	I-H
GOODMAN, Russell, University of New Mexico	V-F
GOTTHELF, Allan, Trenton State University	I-G
GRASSI, Joseph, Fairfield University	I-E
GREENWOOD, John D., City University of New York/City College	V-D
HAACK, Susan, University of Miami	IV-C
HALL, Pamela, Emory University	II-B
HARDING, Sandra, University of Delaware	III-H
HARE, Peter, State University of New York/Buffalo	IV-C
HARRIES, Karsten, Yale University	V-B
HARRIS, Leonard, Purdue University	I-I, III-H
HAWTHORN, John, Syracuse University	V-G
HAYRY, Heta, University of Helsinki, Finland	III-E
HAYRY, Matti, University of Helsinki, Finland	III-E
HEELAN, Patrick, State University of New York/Stony Brook	V-A
HEIL, John, Davidson College, North Carolina	III-F
HENDLEY, Steven, Birmingham Southern College, Alabama	II-E
HIGGINBOTHAM, James, Massachusetts Institute of Technology	V-E

HODGES, Michael, Vanderbilt University	V-D
HOFFMAN, Paul, Massachusetts Institute of Technology	I-F
HOULGATE, Lawrence D.	p. 572
HUNT, David P., Whittier College, California	V-G
HUNTER, Graeme, University of Ottawa, Canada	I-F
HYLAND, Drew, Trinity College, Connecticut	II-D
JACOBSON, Anne, Princeton University	V-D
JALLOH, Chernor, Howard University	I-B
JAMES, Patricia, Kent State University	II-G
JENKINS, John, University of Notre Dame	III-B
JORDAN, Jeff, University of Delaware	I-F
KAZASHI, Nobuo, Yale University	V-F
KEARNS, Thomas, Amherst College, Massachusetts	IV-H
KELLNER, Douglas, University of Texas/Austin	II-J
KELLOGG, Frederick, R., Washington, DC	III-E
KELLY, Michael, Columbia University	II-E
KESTENBAUM, Victor, Boston University	V-F
KISS, Elizabeth, Princeton University	I-E
KITCHER, Philip, University of California/San Diego	V-A
KITTAY, Eva, State University of New York/Stony Brook	IV-A
KIVY, Peter, Rutgers University	II-A
KOHAK, Erazim, Boston University	I-D
KOURANY, Janet, University of Notre Dame	II-B
LADD, Rosalind Ekman, Wheaton College, Massachusetts	II-H
LAND, Berel, State University of New York/Albany	III-C
LANDMAN, Willem, University of the Western Cape, South Africa	II-G
LEVINE, Joseph, North Carolina State University/Raleigh	III-F
LI, Xiaorong, Stanford University	I-J
LINDERMAYER, Eric, Suffolk County Community College, New York	II-I
LIPMAN, Matthew, Montclair State College, New Jersey	II-H
LONG, Roderick, University of North Carolina/Chapel Hill	I-G
LONGINO, Helen, Rice University	II-J, V-A
LYCAN, William G., University of North Carolina/Chapel Hill	I-H
MAGNELL, Thomas, Drew University	III-E
MANDELKER, Steven, East Carolina University	I-H
MANN, William, University of Vermont	I-F
MARCUS, Ruth, Yale University	III-A
MARGOLIS, Joseph, Temple University	III-G
MARKOSIAN, Ned, Lawrence University	V-G
MATTHEN, Mohan, University of Alberta, Canada	IV-B
MATTHEWS, Gareth, University of Massachusetts/Amherst	III-B
MAY, Todd, Clemson University	I-H
MCDERMOTT, John, Texas A & M University	V-B
MCKAY, Thomas J., Syracuse University	V-E
MICHAELS, Meredith, Hampshire College, Massachusetts	II-G
MILLER, Richard, Cornell University	III-D
MINOW, Martha, Harvard Law School	IV-A

MOODY-ADAMS, Michele, Indiana University	I-J
MONTMARQUET, James A., Tennessee State University	I-F
MORGAN, Kathryn, University of Toronto, Canada	II-J
MOSER, Paul, Loyola University of Chicago	V-E
MURRAY, Michael, Franklin and Marshall College, Pennsylvania	I-F
NARAYAN, Uma	p 572
NEVILLE, Robert Cummings, Boston University	IV-I
NUSSBAUM, Martha, Brown University	I-J
NZEGWU, Nkiru, State University of New York/Binghamton	I-J
O'NEILL, Eileen, City University of New York/Queens College and the Graduate Center	III-H
O'SHEA, James, University of North Carolina/Chapel Hill	V-D
OBERDAN, Thomas, Clemson University	V-E
ORMISTON, Gayle, Kent State University	IV-F
OUTLAW, Lucius, Haverford College, Pennsylvania	III-E
PARRY, Richard, Agnes Scott College, Georgia	IV-G
PATTERSON, Dennis M., Rutgers School of Law	IV-H
PELLAUER, David, De Paul University	IV-F
PIKER, Andrew, Kent State University	IV-F
PIPER, Adrian, Wellesley College, Massachusetts	IV-H
PINEAU, Lois, Kansas State University	III-E
PIPPIN, Robert, University of California/San Diego	I-B
POST, John, Vanderbilt University	II-F
POSTOW, Becky, University of Tennessee/Knoxville	III-E
PRESTON, Beth, University of Georgia	III-F
QUICK, James Robert, Duquesne University	III-G
QUINN, Philip, University of Notre Dame	I-F
RESHOTKO, Naomi, University of Denver	IV-G
RING, David, Southern Methodist University	I-F
ROBERTSON, John	p. 572
ROBISON, John, University of Massachusetts/Amherst	V-D
ROCKMORE, Tom, Duquesne University	I-B
ROLLINS, Mark, Washington University	II-F
RORTY, Amelie, Mount Holyoke College, Massachusetts	V-C
ROSENBERG, Jay, University of North Carolina/Chapel Hill	V-B
ROSS, James, University of Pennsylvania	II-C
ROSSMAN, Neil, LaGuardia Community College, New York	II-I
ROTH, Michael, Franklin and Marshall College, Pennsylvania	I-G
ROUSE, Joseph, Wesleyan University	V-A
RUSSON, John, Harvard University	I-G
SARTWELL, Crispin, Vanderbilt University	V-G
SCALTSAS, Patricia Ward, University of Edinburgh, Scotland	II-G
SCALTSAS, Theodore, University of Edinburgh, Scotland	IV-G
SCHAUER, Frederick, Harvard University	I-E
SCHMIDT, Dennis, State University of New York/Binghamton	IV-D
SCHOTT, Robin, University of Louisville	IV-A
SCHRAG, Calvin, Purdue University	III-G
SCHWARTZ, Justin, Ohio State University	II-F

SCHWARZENBACH, Sibyl, City University of New York/ Baruch College	I-G
SCHWEICKART, Patrocinio, University of New Hampshire	II-J
SCOTT, Charles, Vanderbilt University	V-C
SEIGFRIED, Charlene, Purdue University	V-F
SEIGFRIED, Hans, Loyola University of Chicago	V-F
SHAPIRO, Gary, University of Kansas	III-C
SHEN, Vincent, National Chengchih University, China	IV-I
SHER, George, University of Vermont	I-E
SHER, Gila, University of California/San Diego	V-E
SHEROVER, Charles, City University of New York/Hunter College	II-D
SICKELS, Michael David, Pittsburgh, Pennsylvania	III-G
SINGER, Beth, City University of New York/Brooklyn College	IV-E
SKRUPSKELIS, Ignas, University of South Carolina	I-D
SMITH, John-Christian, Youngstown State University	III-F
SMITH, Robin, Kansas State University	I-G
SOSA, Ernest, Brown University	I-D
STUMP, Eleonore, University of Notre Dame	III-B
THOMAS, Laurence, Syracuse University	V-B
THOMSON, Judith J., Massachusetts Institute of Technology	Introduction, Presidential Address
TONG, Lik Kuen, Fairfield University	IV-I
TORRES, Gerald, University of Minnesota	I-I
TRAU, Jane Mary, Barry University	II-G
TUSHNET, Mark, Georgetown Law Center	I-D
UEBEL, Thomas E., Northwestern University	V-E
VAN GULICK, Robert, Syracuse University	III-F
VAN HOOK, Jay M., Northwestern College, Iowa	II-E
VERMA, Roop Rekha, University of Lucknow, India	I-J
VOGEL, Jonathan, Amherst College, Massachusetts	V-G
WAGNER, Steven, University of Illinois/Urbana-Champaign	IV-B
WALKER, Margaret Urban, Fordham University	IV-F
WALLACE, Kathleen, Hofstra University	IV-E
WALSH, Joseph, Stockton State College, New Jersey	III-D
WALTON, Kendall, University of Michigan	I-C, II-A
WARD, Andrew, San Jose State University	V-D
WARTOFSKY, Marx, City University of New York/Baruch College	III-C
WEISSMAN, David, City University of New York/City College	I-A
WELCHMAN, Jennifer, Colgate University	V-F
WESTPHAL, Merold, Fordham University	II-B
WILLIAMS, Patricia, University of Wisconsin	I-I
WILSHIRE, Bruce, Rutgers University	V-F
WITT, Charlotte, University of New Hampshire	I-G
WOLFF, Robert Paul, University of Massachusetts/Amherst	I-E
WOLTERSTORFF, Nicholas, Yale University	I-C
WRIGHT, Kathleen, Haverford College, Pennsylvania	IV-D
WYSCHOGRAD, Edith, City University of New York/Queens College	II-D

APPENDIX: PROGRAMS OF GROUP MEETINGS

FRIDAY, DECEMBER 27

I. EVENING (8:00–11:00)

1. 8:00 LENSGRINDER'S SOCIETY, *Odets/Wilder*
Reception for Job Seekers

2. 8:00 SOCIETY FOR THE STUDY OF PROCESS PHILOSOPHIES *Kern/Sullivan*
Topic: George Lucas's *The Rehabilitation of Whitehead*
Chair: Joseph Grange
Speakers: Donald Sherburne and David Kite
Respondent: George Lucas

3. 8:00 SOCIETY FOR ASIAN AND COMPARATIVE PHILOSOPHY *Juliard/Imperial*
Topic: Culture and Self: Asian Perspectives
Chair: Ashok Malhotra
I. Purusottima Bilimoria, "Classical Indian Conception of the Self"
II. Sam Fohr, "Self as a Reflection of Culture in Hindu and Western Society"
III. Douglas Allen, "Indian, Marxist, and Feminist Critiques of the Self"
IV. Alex Wayman, "No-self of Buddhism within Indian Culture"
Commentators: Jeffrey Timm, Hope Fitz, A.N. Balslev, and Frank Hoffman

4. 8:00 PERSONALIST DISCUSSION GROUP, *Broadhurst/Belasco*
Chair: Thomas O. Buford
John Lachs, "Persons and Different Sorts of Persons"

5. 8:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Schubert/Uris*
Topic: Hegel and the Theory of Reference
Chair: Rona Burger
Speaker: Katharina Dulckeit
Commentators: Arindam Chakrabarti and Robert B. Berman

6. 8:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Carnegie/Lyceum*
Chair: John A. Loughney
Board of Officers' Annual Business Meeting

SATURDAY, DECEMBER 28

II. MORNING (9:00–11:00)

1. 9:00 SOCIÉTÉ AMÉRICAINE DE PHILOSOPHIE DE LANGUE FRANÇAISE *O'Neill*
Chair: Peter Caws
I. Tom Rockmore, "The Reception of Heidegger in Recent French Thought"

II. Hugh Silverman, "Difference and the Differend—Derrida and Lyotard"
III. David Williams, "Condorcet and the English Enlightenment"

2. 9:00 RADICAL PHILOSOPHY ASSOCIATION, *Odets/Wilder*
Topic: Alan Gilbert's *Democratic Individuality*
Chair: Justin Schwartz
Speakers: Andrew Levine and Richard Farr
Respondent: Alan Gilbert

3. 9:00 AMERICAN SOCIETY FOR VALUE INQUIRY, *Miller*
Chair: Ruth Lucier
John M. Abbarno, "Role Responsibility and Values" (Presidential Address)
Business Meeting

4. 9:00 GANDHI-KING SOCIETY, *Albee*

5. 9:00 WORLD PHENOMENOLOGY INSTITUTE, *Hart*
I. Anna-Teresa Tymieniecka, "Faith and Belief"
II. Terry S. Kasely, "Faith and Intellectual Fortitude"

6. 9:00 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, *Brecht*
Topic: Animals, Caring, and the Status of Infants
Chair: Lilly-Marlene Russow
I. Brian Luke, "Justice, Caring, and Animal Liberation"
Commentator: Timothy Menta
II. Stephen G. Post, "Animal Rights Theory and the Diminishment of Infants"
Commentator: Harlan B. Miller

7. 9:00 SOCIETY FOR PHILOSOPHY AND TECHNOLOGY, *Plymouth/Royale*
Topic: In the Current Technoeconomic Situation, Is Socialism Still Viable?
Chair: Paul T. Durbin
Andrew Feenberg, "A Critical Theory of Technology"
Commentator: Robert S. Cohen

8. 9:00 SOCIETY FOR THE PHILOSOPHIC STUDY OF THE CONTEMPORARY VISUAL ARTS, *Carnegie/Lyceum*
I. Kendall D'Andrade, "Who Is the Man for All Seasons?"
II. J. Stanley Yake, "Alice, Plato, and the Ring of Truth"

III. LATE MORNING (11:15–1:15)

1. 11:15 SOCIETY FOR INDIAN PHILOSOPHY AND RELIGION, *Gilbert*
Speakers: Mark Siderits and Steven Laycock
Commentators: Kisor Chakrabarti and Chandana Chakrabarti

2. 11:15 ASSOCIATION FOR PHILOSOPHY AND LIBERATION, *O'Neill*
Topic: Liberation Philosophy and Theology in North America and Latin America
Chair: Michael Barber

Saturday, December 28, Late Morning, III-2

I. Eduardo Mendieta, "Theology of Liberation and Liberation Philosophy: A Latin American Perspective"
II. Kate Lindemann, "Philosophy of Liberation in a North American Context"

3. 11:15 SOCIETY FOR THE ADVANCEMENT OF AMERICAN PHILOSOPHY *Odets/Wilder*
Topic: Charlene Haddock Seigfried's *William James's Radical Reconstruction of Philosophy*
Chair: William J. Gavin
I. Vincent Colapietro, "James's Deconstruction of Reason: Otherness, Openness, and Fluency"
II. Eugene Fontinell, "James's Reconstruction of Religion: Pragmatic—Processive—Pluralistic"
Respondent: Charlene Haddock Seigfried

4. 11:15 SØREN KIERKEGAARD SOCIETY, *Miller*
Chair: Merold Westphal
I. Robert R. Williams, "Morality and God: Kierkegaard's Teleological Suspension of the Ethical"
II. Martin J. Matustik, "Habermas's Reading of Kierkegaard: The Possibility of a Communicative and Existential Ethic"

5. 11:15 SARTRE CIRCLE, *Albee*
Chair: Phyllis S. Morris
I. Debra Bergoffen, "The Look as Bad Faith: Sartre and de Beauvoir"
Commentator: Eleanore Holveck
II. David Jopling, "Sartre and Cognitive Psychology"
Commentator: Kathleen Wider

6. 11:15 SOCIETY OF PHILOSOPHERS AT WORK IN THE WORLD, *Hart*
Topic: What Does Philosophy Have to Do With Business?
Speakers: Edward Freeman and Stephanie R. Lewis

7. 11:15 SOCIETY FOR THE PHILOSOPHICAL STUDY OF MARXISM, *Brecht*
Topic: The State of the Socialist Idea in the USSR Today
Chair: Edward D'Angelo
Speakers: Sarah Fletcher Luther, John J. Neumaier, and Howard L. Parsons

8. 11:15 SOCIETY FOR WOMEN IN PHILOSOPHY, *Schubert/Uris*
Topic: Feminism, Desire, and the Body (In Honor of Linda Singer)
Chair: Kelly Oliver
Speakers: Judith Butler, Iris Young, and Tina Chanter

9. 11:15 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Kern/Sullivan*
Chair: David Stern
I. Kenley R. Dove, "The Global Dimension of Ethical Life"
II. Kenneth Smith, "Consciousness and *Verstand:* Modernity as Ideology"

IV. AFTERNOON (2:00–5:00)

1. 2:00 SOCIETY FOR BUSINESS ETHICS, *O'Neill*
Topic: Business Ethics and Legal Philosophy
Chair: Jennifer Moore
I. Alan Wertheimer, "Unconscionability and Contracts"
II. Leo Katz, "The Extortionate Bargain"

2. 2:00 INTERNATIONAL SOCIETY FOR CHINESE PHILOSOPHY, *Miller*
Chair: Kwang-Sae Lee
I. Michael K. Green, "Culture, Self, and Ethical Paradigms"
II. Kwong-Loi Shun, "Moral Motivation in Confucian Ethics"
III. Kwang-Sae Lee, "Two Ways of Virtue: Confucian and Kantian"

3. 2:00 INTERNATIONAL SOCIETY FOR ENVIRONMENTAL ETHICS, *Albee*
Chair: Eric Katz
I. Gary E. Varner, "A Critique of Environmental Holism"
Commentator: Peter S. Wenz
I. David Abram, "On the Ecological Consequences of Alphabetical Literacy"
Commentator: Bruce Morito

4. 2:00 SOCIETY FOR IBERIAN AND LATIN AMERICAN THOUGHT, *Hart*
Topic: Ethics and Conquest: Tzvetan Todorov's *The Conquest of America*
Speakers: Oscar Marti, Amy Oliver, Andrés Salgado, Ofelia Schutte, and Fred Sturm
Business Meeting

V. LATE AFTERNOON (5:15–7:15)

1. 5:15 CHARLES S. PIERCE SOCIETY, *O'Neill*
Chair: Michael Shapiro
I. Robert W. Burch, "Game Theoretical Semantics for Peirce's Existential Graphs"
II. John J. Fitzgerald, "Peirce On Icons"
III. Peirce Essay Contest Winner
Business Meeting

2. 5:15 SOCIETY FOR ANCIENT GREEK PHILOSOPHY, *Odets/Wilder*
Topic: Plato
Chair: John P. Anton
I. Reginald Allen, "The Proof of Immortality from Self-moving Motion in Plato's *Phaedrus*"
II. Allan Silverman, "Some Ways of Being in Plato"

3. 5:15 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, *Albee*

4. 5:15 SOCIETY FOR MACHINES AND MENTALITY, *Kern/Sullivan*
Chair: William J. Rapaport
William Bechtel, "Currents in Connectionism"
Reception in Honor of *Minds and Machines* (Sponsored by Kluwer Academic Publishers, Chair: James H. Fetzer)

Saturday, December 28, Late Afternoon, V

5. 5:15 SOCIETY FOR PHILOSOPHY AND PUBLIC AFFAIRS, *Juliard/Imperial*
Topic: Feminist Jurisprudence and the Legal Subordination of Women
Chair: Rosemarie Tong
I. Patricia G. Smith, "Feminist Jurisprudence and the Nature of Law"
II. Robin West, "Reconstructing Liberty"
III. Joan Williams, "Is Polygamy Good Feminism?"

6. 5:15 SOCIETY FOR THE STUDY OF HUSSERL'S PHILOSOPHY *Broadhurst/Belasco*
Topic: Derrida on Husserl
Chair: J. N. Mohanty
Speakers: William McKenna and Claude Evans

7. 5:15 JASPERS SOCIETY OF NORTH AMERICA, *Plymouth/Royale*
Topic: Heidegger and Jaspers
Chair: George Pepper
I. Krystyna Gorniak-Kocikowska, "The Concept of Freedom in Jaspers and Heidegger"
Commentator: Joseph Margolis
II. Harold H. Oliver, "The Psychological Dimension in Jaspers's Relationship with Heidegger"
Commentator: Tom Rockmore

8. 5:15 ASSOCIATION FOR THE PHILOSOPHY OF THE UNCONSCIOUS *Palace/Winter Garden*
Topic: The Order of the Real: Nietzsche and Lacan
Chair: Wilfried Ver Eecke
Speaker: Babette E. Babich
Commentator: Debra Bergoffen

9. 5:15 ASSOCIATION OF PHILOSOPHY JOURNAL EDITORS, *Miller*
Reception for Present and Prospective Members

VI. EVENING (7:30–10:30)

1. 7:30 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING *O'Neill*
Chair: J. Anthony Blair
I. John Woods
Commentators: L. Jonathan Cohen and William G. Lycan
Chair: Mark Weinstein
II Lenore Langsdorf, "Rhetoric and Informal Logic/Critical Thinking: Are There Differences That Make a Difference?" (Presidential Address)

2. 7:30 SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, *Odets/Wilder*
Chair and Commentator: Edith Wyschograd
I. James Watson, "Why Heidegger Was Not Shocked by the Holocaust"
II. Alan Milchman and Alan Rosenberg, "The Holocaust in Light of Heidegger's Understanding of Technology"
III. John D. Caputo, "Hegel's Scandal"

3. 7:30 INTERNATIONAL BERKELEY SOCIETY, *Miller*
Topic: Berkeley's Philosophy of Science
Chair: Kenneth Winkler
I. Lisa Downing, "Berkeley's Instrumentalism about Dynamics"
II. Catherine Wilson, "Berkeley and the Microworld"
Commentator: Richard Brook

4. 7:30 INTERNATIONAL SOCIETY FOR UNIVERSALISM, *Albee*
Topic: Individualism, Communalism, and Univeralism in Hegel's Ethical Thought
Chair: Michael H. Mitias
I. Robert McKeon, "The Ethics of the Free Market: Norms for a Just World Order"
Chair: O. William Rhodenhiser
II. Albert A. Anderson, "Universal Justice: The Alternative to Modernism"
Chair: Charles S. Brown
III. Steven Hicks, "Individualism, Communalism, and Universalism in Hegel's Ethical Thought"

5. 7:30 NORTH AMERICAN KANT SOCIETY, *Hart*
Topic: Rudolf A. Makkreel's ***Imagination and Interpretation in Kant***
Speakers: David Carr and Guenter Zoeller
Respondent: Rudolf A. Makkreel

6. 7:30 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY *Kern/Sullivan*
Chair: Edward Mahoney
I. Paul Oscar Kristeller, "Napoleonic Influences in Medieval and Renaissance Philosophy"
II. Herbert Davidson, "Averroes, A Philosopher Galloping Off in Opposite Directions"

7. 7:30 NORTH AMERICAN SOCIETY FOR SOCIAL PHILOSOPHY, *Brecht*
Topic: Market Socialism
Chair: Michael W. Howard
I. David Schweickart, "Economic Democracy: Worthy Socialism that Would Really Work"
II. N. Scott Arnold, "Exploitation in the Market Socialist Economy"

Saturday, December 28, Evening, VI-7

III. David Gilboa, "Where Have All the Profits Gone? Arnold on the Degeneration of Market Socialism"
Commentator: Justin Schwartz

8. 7:30 AMERICAN SOCIETY FOR VALUE INQUIRY AND THE INTERNATIONAL SOCIETY FOR ENVIRONMENTAL ETHICS, *Juliard/Imperial*
Topic: Advocacy and Values
Co-Chairs: Laura Westra and John M. Abbarno
I. Tom Regan, "The Proper Business of the Moral Philosopher"
II. Kristin Shrader-Frechette, "Ethical Advocacy and Environmental Values"
Commentators: Robert Fullinwinder and William Aiken

9. 7:30 ASSOCIATION FOR PHILOSOPHY OF EDUCATION *Broadhurst/Belasco*
Topic: The Moral Basis of Liberal Education
Chair: Sophie Haroutunian-Gordon
Speaker: Alan Gewirth

10. 7:30 SOCIETY FOR WOMEN IN PHILOSOPHY, *Carnegie/Lyceum*
Topic: In Honor of Distinguished Women Philosophers
Chair: Patrice DiQuinzio
Jane Roland Martin (1991 Honoree), "Essentialism, False Difference, and Other Dangerous Traps"
Business Meeting
Chair: Margaret Walsh

11. 7:30 SOCIETY FOR ASIAN AND COMPARATIVE PHILOSOPHY, *Salon A*
Topic: Culture and Self: Comparative Perspectives
Chair: John Koller
I. Alan Roland, "Self in India, Japan, and the West"
II. Henry Rosemont, "Why Role-bearing Persons Cannot Be Rights-bearing Individuals"
III. W.G. Frisina, "Neo-Pragmatism and Neo-Confucianism"
IV. Graham Parkes, "Nietzsche and Nishitani on the Postnihilism Relation to Tradition"
Commentators: Russell Blackwood, Mike Barnhart, John Tucker, and Yutaka Yamamoto

12. 7:30 LEIBNIZ SOCIETY OF NORTH AMERICA, *Salon B/C*
Chair: Robert C. Sleigh, Jr.
Eileen A. O'Neill, "The Role of Unintelligibility Arguments in the Leibniz/Masham, Descartes/Elisabeth Correspondences"
Business Meeting

13. 7:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, *Cantor/Jolson*
Topic: The Role of Philosophy in American Higher Education
Chair: John A. Loughney
Speakers: David Weissman, Nancy Simco, John A. Loughney, and John E. Smith

14. 7:30 CONCERNED PHILOSOPHERS FOR PEACE, *Gilbert*
Topic: Just War in the Persian Gulf?
Chair: Barry Gan
I. Robert L. Holmes, "War and Morality: The Persian Gulf War"
Commentator: Joseph Betz
II. Jonathan Schonscheck, "Political Realism, Morality, and the Gulf War"
Commentator: Richard Werner
Business Meeting

SUNDAY, DECEMBER 29

VII. MORNING (9:00–11:00)

1. 9:00 SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, *O'Neill*
Topic: Berel Lang's *Act and Idea in the Nazi Genocide*
Chair: Alan Rosenberg
Speakers: Laurence Thomas and Garry Brodsky
Respondent: Berel Lang
Business Meeting

2. 9:00 HUME SOCIETY, *Odets/Wilder*
Topic: Hume on Religion
Chair: Geoffrey Sayre-McCord
I. Annette Baier, "Hume's Reflective Religious Sentiments"
II. Don Garrett, "Divine Causation and Human Judgment"

3. 9:00 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY, *Albee*
Topic: David L. Norton's *Democracy and Moral Development*
Chair: John Hasnas
Speakers: Matthew Freytag and Douglas Rasmussen
Respondent: David L. Norton

4. 9:00 WORLD PHENOMENOLOGY INSTITUTE, *Brecht*
I. Steven W. Laycock, "Oneness and Emptiness: A Study of Dumery and Buddhism"
II. Marlies Kronegger, "Faith, Virtues, and Passions in Baroque Literature"

5. 9:00 SOCIETY FOR MEDIEVAL AND RENAISSANCE PHILOSOPHY *Juliard/Imperial*
Chair: Stephen Menn
I. Deborah Black, "Practical Wisdom and Moral Virtue in Arabic Philosophy"
II. Marilyn Adams, "Anselm's Method in Philosophical Theology"

Sunday, December 29, Morning, VII

6 9:00 SOCIETY FOR THE PHILOSOPHIC STUDY OF THE CONTEMPORARY VISUAL ARTS, *Hart*
Chair: Kendall D'Andrade
I. Graham Parkes, "Phantasy Projections of the Multiple Self in the Films of Bergman, Fellini, and Resnais"
II. Noël Carroll, "The Nature of Mass Art"

VIII. AFTERNOON (2:15–4:15)

1. 2:15 AYN RAND SOCIETY, *O'Neill*
Topic: Ayn Rand and Kant
Chair: Allan Gotthelf
Speaker: George V. Walsh
Commentator: Fred D. Miller, Jr.

2. 2:15 SOCIETY FOR THE PHILOSOPHICAL STUDY OF MARXISM, *Miller*
Topic: Socialism, Art, and Aesthetics
Chair: John Ryder
Speakers: David Craven and Dimitri Khanin

3. 2:15 SOCIETY FOR THE PHILOSOPHY OF CREATIVITY, *Albee*
Topic: Pierre Duhem and Creativity
Chair: Jennifer McErlean
Anastasios A. Brenner, "Duhem's Energetics and Philosophical Creativity"
Commentator: Roger Ariew

4. 2:15 RADICAL PHILOSOPHY ASSOCIATION, *Hart*
Topic: Homelessness, Poverty, and the Law
Chair/Commentator: Patricia Mann
I. Uma Narayan, "Is the Ban on Begging Justified?"
II. Steven Jay Gold, "Police Brutality"

5. 2:15 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Juliard/Imperial*
Topic: Social Justice and the Limits of the Market
Chair: Andrew Buchwalter
Speaker: David P. Levine
Commentators: Tony Smith and Richard D. Winfield

6. 2:15 NORTH AMERICAN NIETZSCHE SOCIETY, *Kern/Sullivan*
Topic: Nietzsche on Virtue and Ethics
Chair: Alexander Nehamas
Speakers: Lester Hunt, Martha Nussbaum, Kelly Oliver, and Alan White

IX. EVENING (6:30–9:30)

1. 6:30 SOCIETY FOR REALIST/ANTIREALIST DISCUSSION, *Gilbert*
Chair: Mary Carman Rose
I. Peter Ochs, "Pragmatism and Theosemiotics"
II. Mary Carman Rose, "Contemporary Issues of Realism and the Theism of Peirce"

2. 6:30 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING *O'Neill*

Topic: Social Epistemology
Chair: Helen Longino
Speakers: Steven Fuller and Alvin I. Goldman
Commentator: Philip Kitcher
Business Meeting

3. 6:30 INTERNATIONAL SOCIETY FOR ENVIRONMENTAL ETHICS, SOCIETY FOR THE PHILOSOPHIC STUDY OF GENOCIDE AND THE HOLOCAUST, AND RADICAL PHILOSOPHY ASSOCIATION *Odets/Wilder*

Topic: Holocaust, Genocide, and Ecocide
Chair: Bat-Ami Bar-On
I. Eric Katz, "The Death of Nature"
II. Alan Rosenberg, "The Left and the Holocaust"
III. Roger S. Gottlieb, "How Can We Face the Truth?"

4. 6:30 SOCIETY FOR LESBIAN AND GAY PHILOSOPHY, *Miller*
Topic: The Ethics of Outing: Arguments for and against
Chair: Noretta Koertge
I. Mark Chekola, "Outing: Cases and Principles"
II. David Mayo, "Privacy, Hypocrisy, and the Ethics of Outing"
Commentator: Jerry McCarthy

5. 6:30 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, *Albee*
Chair: Harlan B. Miller
I. Gary Comstock, "Should We Genetically Engineer Hogs?"
Chair: David J. Detmer
II. Natalie Dandekar and Edward Zlotkowski, "Moral Issues Associated with Bioengineered Species—Stewardship, Abuse, and Sustainability"
Commentator: Laura Westra

6. 6:30 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, *Hart*

7. 6:30 SOCIETY FOR THE PHILOSOPHY OF SEX AND LOVE, *Kern/Sullivan*
Topic: Pornography, Art, and the Origins of Consciousness
Chair: Rosemarie Tong
Speaker: Justin Leiber
Commentator: Lenore Langsdorf

8. 6:30 JASPERS SOCIETY OF NORTH AMERICA, *Juliard/Imperial*
Topic: The University and the Civil Society
Chair: Alan M. Olson
I. Robert C. Neville, "The University as Moral and Spiritual Prism"
Commentator: Charles L. Griswold, Jr.
II. Bruce Wilshire, "The Moral Collapse of the University: Professionalism, Purity, and Alienation"
Commentator: Patricia Mann

Sunday, December 29, Evening, IX

9. 6:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES *Broadhurst/Belasco*

Topic: Philosophy, Counseling, and Psychotherapy

Chair: Thomas Magnell

I. Brad Art, "Chuang Tsu's Butterfly and Other Confusing Encounters"

II. Elliot Cohen, "The Virtuous Therapist"

III. Harold L. Parker, "Disease or Not Disease, Is that the Question? Confusions in the Language of Addiction Counseling"

IV. Paul Sharkey, "Philosophy, Science, and Quackery"

Business Meeting: American Society for Philosophy, Counseling, and Psychotherapy

MONDAY, DECEMBER 30

X. MORNING (9:00–11:00)

1. 9:00 ASSOCIATION FOR PHILOSOPHY OF EDUCATION, *Odets/Wilder*

Topic: Reforming the Curriculum: Traditionalist, Multicultural, and Radical Pedagogies

Chair: Emily Robertson

Panel: Christopher L. Pines, Alven Neiman, Steven Schroeder, and Alan Gewirth

Business Meeting

2. 9:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, *Kern/Sullivan*

Topic: Skepticism and Systematic Philosophy

Chair: James Wilkinson

Speakers: Donald Livingston and Chip Sills

3. 9:00 INTERNATIONAL SOCIETY FOR NEOPLATONIC STUDIES, *Hart*

4. 9:00 WORLD PHENOMENOLOGY INSTITUTE, *Brecht*

I. Thomas Ryba, "Magister Internis: A Phenomenology of Faith as Desire and Teacher"

II. Krystyna Gorniak-Kocikowska, "Approaches to Faith and Virtue"

XI. LATE MORNING (11:15–1:15)

1. 11:15 RADICAL PHILOSOPHY ASSOCIATION, *Miller*

Topic: Postmodern Dialogues

Chair: Rodney Peffer

I. Richard T. Peterson, "Dialogue Versus Dialectic: Bakhtin and Political Criticism"

II. Isaac Nevo, "Rorty's Romantic Pragmatism"

III. Bill Martin, "Against the Strategy of Cynicism: Radical Politics in the Margins of the Postmodern Condition"

2. 11:15 SARTRE CIRCLE, *Albee*
   Chair: Robert V. Stone
   I. Bruce Baugh, "The Unhappy Consciousness in Sartre's *Being and Nothingness*"
   Commentator: Frithjof Bergmann
   II. Raymond Langley, "The Kantian Dimensions of Sartre's Late Projects"
   Commentator: Ronald Aronson

3. 11:15 SOCIETY FOR WOMEN IN PHILOSOPHY, *Carnegie/Lyceum*
   Topic: Colonialism and Philosophical Exclusion: Toward an Understanding of Constructive Dialogue
   Chair: Susan Bernick
   I. Nkiru Nzegwu, "Construction of Female Political Identity in a Dual-sex System"
   II. Jackie Davies, "A Different Ear (for White Women): Learning To Hear Native Women's Voices"
   III. Susan Babbitt, "Pushing the Limits: Engagement and the Possibility of Radical Ideological Critique"

4. 11:15 SOCIETY FOR REALIST/ANTIREALIST DISCUSSION, *O'Neill*
   Chair: Mary Carman Rose
   I. Guy S. Axtell, "Pragmatism and Dialectical Epistemology"
   II. Dennis Patterson, "The Legal Case against Realism"

GROUP PARTICIPANTS

(See Appendix)

ABBARNO, John M., D'Youville College, New York	II-3, VI-8
ABRAM, David, State University of New York/Stony Brook	IV-3
ADAMS, Marilyn, University of California/Los Angeles	VII-5
AIKEN, William, Chatham College, Pennsylvania	VI-8
ALLEN, Douglas, University of Maine/Orono	I-3
ALLEN, Reginald, Northwestern University	V-2
ANDERSON, Albert A., Babson College, Massachusetts	VI-4
ANTON, John P., University of South Florida/Tampa	V-2
ARIEW, Roger, Virginia Polytechnic Institute and State University	VIII-3
ARNOLD, N Scott, University of Alabama/Birmingham	VI-7
ARONSON, Ronald, Wayne State University/Weekend College	XI-2
ART, Brad, Westfield State College, Massachusetts	IX-9
AXTELL, Guy S., University of Hawaii/Manoa	XI-4
BABBITT, Susan, Queens University, Canada	XI-3
BABICH, Babette E , Fordham University	V-8
BAIER, Annette, University of Pittsburgh	VII-2
BALSLEV, A. N.	I-3
BARBER, Michael, St. Louis University	III-2

BARNHART, Mike, Temple University VI-11
BAR-ON, Bat-Ami, State University of New York/Binghamton IX-3
BAUGH, Bruce, University of Victoria, Canada XI-2
BECHTEL, William, Georgia State University V-4
BERGMANN, Frithjof, University of Michigan/Ann Arbor XI-2
BERGOFFEN, Debra, George Mason University III-5, V-8
BERMAN, Robert B., Xavier University I-5
BERNICK, Susan, Furman University XI-3
BETZ, Joseph, Villanova University VI-14
BILIMORIA, Purusottima, Deakin University, Australia I-3
BLACK, Deborah, University of Toronto, Canada VII-5
BLACKWOOD, Russell, Hamilton College, New York VI-11
BLAIR, J. Anthony, University of Windsor, Canada VI-1
BRENNER, Anastasios A., Université de Technologie de Compiegne, France VIII-3
BRODSKY, Garry, University of Connecticut VII-1
BROOK, Richard, Bloomsburg University VI-3
BROWN, Charles S., Emporia State University VI-4
BUCHWALTER, Andrew, University of North Florida VIII-5
BUFORD, Thomas O., Furman University I-4
BURCH, Robert, W., Texas A & M University V-1
BURGER, Rona, Tulane University I-5
BUTLER, Judith, Johns Hopkins University III-8
CAPUTO, John D., Villanova University VI-2
CARR, David, University of Ottawa, Canada VI-5
CARROLL, Noël, Cornell University VII-6
CAWS, Peter, George Washington University II-1
CHAKRABARTI, Arindam, University of Washington I-5
CHAKRABARTI, Chandana, Elon College, North Carolina III-1
CHAKRABARTI, Kisor, University of New Orleans III-1
CHANTER, Tina, Louisiana State University III-8
CHEKOLA, Mark, Moorehead State University IX-4
COHEN, Elliot, Indian River Community College, Florida IX-9
COHEN, L. Jonathan, Oxford University, England VI-1
COHEN, Robert S., Boston University II-7
COLAPIETRO, Vincent, Fordam University III-3
COMSTOCK, Gary, Iowa State University IX-5
CRAVEN, David, State University of New York/Cortland VIII-2
DANDEKAR, Natalie, University of Rhode Island IX-5
DAVIDSON, Herbert, University of California/Los Angeles VI-6
DAVIES, Jackie, Queens University, Canada XI-3
DETMER, David J., Purdue University/Calumet IX-5
DIQUINZIO, Patrice, University of Scranton VI-10
DOVE, Kenley R., State University of New York/Purchase III-9
DOWNING, Lisa, Yale University VI-3
DULCKEIT, Katharina, Butler University I-5
DURBIN, Paul T., University of Delaware II-7
D'ANDRADE, Kendall, Shawnee State University II-8, VII-6
D'ANGELO, Edward, University of Bridgeport III-7
EVANS, Claude, Washington University V-6
FARR, Richard, University of Hawaii/Manoa II-2

FEENBERG, Andrew, San Diego State University	II-7
FETZER, James H., University of Minnesota/Duluth	V-4
FITZ, Hope, Eastern Connecticut State University	I-3
FITZGERALD, John J., Southeastern Massachusetts University	V-1
FOHR, Sam, University of Pittsburg/Bradford	I-3
FONTINELL, Eugene, City University of New York/Queens College	III-3
FREEMAN, Edward, University of Virginia	III-6
FREYTAG, Matthew, University of Richmond	VII-3
FRISINA, W. G., University of Houston	VI-11
FULLER, Steven, Virginia Polytechnic Institute and State University	IX-2
FULLINWINDER, Robert, University of Maryland	VI-8
GAN, Barry, St. Bonaventure University	VI-14
GARRETT, Don, University of Utah	VII-2
GAVIN, William J., University of Southern Maine	III-3
GEWIRTH, Alan, University of Chicago	VI-9, X-1
GILBERT, Alan, University of Denver	II-2
GILBOA, David, Berea College, Kentucky	VI-7
GOLD, Steven Jay, Southern Connecticut State University	VIII-4
GOLDMAN, Alvin I., University of Arizona	IX-2
GORNIAK-KOCIKOWSKA, Krystyna, Temple University and University of Rhode Island	V-7, X-4
GOTTHELF, Allan, Trenton State College, New Jersey	VIII-1
GOTTLIEB, Roger S., Worchester Polytechnic Institute	IX-3
GRANGE, Joseph, University of Southern Maine	I-2
GREEN, Michael K., State University of New York/Oneonta	IV-2
GRISWOLD, JR., Charles L., Boston University	IX-8
HAROUTUNIAN-GORDON, Sophie, University of Chicago	VI-9
HASNAS, John, Georgetown University	VII-3
HICKS, Steven, Concordia College, Minnesota	VI-4
HOFFMAN, Frank, West Chester University	I-3
HOLMES, Robert L., University of Rochester	VI-14
HOLVECK, Eleanore, Duquesne University	III-5
HOWARD, Michael W., University of Maine/Orono	VI-7
HUNT, Lester, University of Wisconsin/Madison	VIII-6
JOPLING, David, Emory University	III-5
KASELY, Terry S., Indiana University of Pennsylvania	II-5
KATZ, Eric, New Jersey Institute of Technology	IV-3, IX-3
KATZ, Leo, University of Pennsylvania	IV-1
KHANIN, Dimitri, Colgate University	VIII-2
KITCHER, Philip, University of California/San Diego	IX-2
KITE, David, Emory University	I-2
KOERTGE, Noretta, Indiana University	IX-4
KOLLER, John, Rensselaer Polytechnic Institute	VI-11
KRISTELLER, Paul Oscar, Columbia University	VI-6
KRONEGGER, Marlies, Michigan State University	VII-4
LACHS, John, Vanderbilt University	I-4
LANG, Berel, State University of New York/Albany	VII-1
LANGLEY, Raymond, Manhattanville College, New York	XI-2
LANGSDORF, Lenore, Southern Illinois University/Carbondale	VI-1, IX-7
LAYCOCK, Steven, University of Toledo	III-1, VII-4

LEE, Kwang-Sae, Kent State University IV-2
LEIBER, Justin, University of Houston IX-7
LEVINE, Andrew, University of Wisconsin/Madison II-2
LEVINE, David P., University of Denver VIII-5
LEWIS, Stephanie R., Princeton, New Jersey III-6
LINDEMANN, Kate, Mt. St. Mary College, New York III-2
LIVINGSTON, Donald, Emory University X-2
LONGINO, Helen, Rice University IX-2
LOUGHNEY, John A., Westfield State College, Massachusetts I-6, VI-13
LUCAS, George, Clemson University I-2
LUCIER, Ruth, Bennet College, North Carolina II-3
LUKE, Brian, University of Pittsburgh II-6
LUTHER, Sarah Fletcher III-7
LYCAN, William G., University of North Carolina/Chapel Hill VI-1
MAGNELL, Thomas, Drew University IX-9
MAHONEY, Edward, Duke University VI-6
MAKKREEL, Rudolf A., Emory University VI-5
MALHOTRA, Ashok, State University of New York/Oneonta I-3
MANN, Patricia, City University of New York/City College VIII-4, IX-8
MARGOLIS, Joseph, Temple University V-7
MARTI, Oscar, University of California/Los Angeles IV-4
MARTIN, Bill, DePaul University IX-1
MARTIN, Jane Roland, University of Massachusetts/Boston VI-10
MATUSTIK, Martin J., Purdue University III-4
MAYO, David, University of Minnesota/Duluth IX-4
MCCARTHY, Jerry, City University of New York/College of Staten Island IX-4
MCERLEAN, Jennifer, Siena College, New York VIII-3
MCKENNA, William, Miami University of Ohio V-6
MCKEON, Robert, Babson College, Massachusetts VI-4
MENDIETA, Eduardo, Union Theological Seminary, New York III-2
MENN, Stephen, Princeton University VII-5
MENTA, Timothy, Boston University II-6
MILCHMAN, Alan, City University of New York/Queens College VI-2
MILLER, Harlan B., Virginia Polytechnic Institute and State University II-6, IX-5
MILLER, Jr., Fred D., Bowling Green State University VIII-1
MITIAS, Michael H., Millsaps College, Mississippi VI-4
MOHANTY, J. N., Temple University V-6
MOORE, Jennifer, University of Delaware IV-1
MORITO, Bruce, University of Guelph, Canada IV-3
MORRIS, Phyllis S., Ann Arbor, Michigan III-5
NARAYAN, Uma, Vassar College, New York VIII-4
NEHAMAS, Alexander, Princeton University VIII-6
NEIMAN, Alven, University of Notre Dame X-1
NEUMAIER, John J., Poughkeepsie, New York III-7
NEVILLE, Robert C., Boston University IX-8
NEVO, Isaac, Ben Gurion University of the Negev, Israel XI-1
NORTON, David L., University of Delaware VII-3
NUSSBAUM, Martha, Brown University VIII-6
NZEGWU, Nkiru, State University of New York/Binghamton XI-3
OCHS, Peter, Drew University IX-1

OLIVER, Amy, American University IV-4
OLIVER, Harold H , Boston University V-7
OLIVER, Kelly, Miami University of Ohio III-8, VIII-6
OLSON, Alan M., Boston University IX-8
O'NEILL, Eileen A , City University of New York/Queens College and the Graduate Center VI-12
PARKER, Harold L., Middle Tennessee State University IX-9
PARKES, Graham, University of Hawaii/Manoa VI-11, VII-6
PARSONS, Howard L , University of Bridgeport III-7
PATTERSON, Dennis, Rutgers University XI-4
PEFFER, Rodney, University of San Diego XI-1
PEPPER, George, Iona College, New York V-7
PETERSON, Richard T., Michigan State University XI-1
PINES, Christopher L., University of Rio Grande X-1
POST, Stephen G., Case Western Reserve University II-6
RAPAPORT, William J., State University of New York/Buffalo V-4
RASMUSSEN, Douglas, St. John's University VII-3
REGAN, Tom, North Carolina State University VI-8
RHODENHISER, O. William, University of Richmond VI-4
ROBERTSON, Emily, Syracuse University X-1
ROCKMORE, Tom, Duquesne University II-1, V-7
ROLAND, Alan, New York, New York VI-11
ROSE, Mary Carman, Goucher College, Maryland IX-1, XI-4
ROSEMENT, Henry, St Mary's College, Maryland VI-11
ROSENBERG, Alan, City University of New York/Queens College VI-2, VII-1, IX-3
RUSSOW, Lilly-Marlene, Purdue University II-6
RYBA, Thomas, Purdue University X-4
RYDER, John, State University of New York/Cortland VIII-2
SALGADO, Andrés, Catholic University of Louvain/Belgium IV-4
SAYRE-MCCORD, Geoffrey, University of North Carolina/Chapel Hill VII-2
SCHONSHECK, Jonathan, Le Moyne College, New York VI-14
SCHROEDER, Steven, Roosevelt University X-1
SCHUTTE, Ofelia, University of Florida IV-4
SCHWARTZ, Justin, Ohio State University II-2, VI-7
SCHWEICKART, David, Loyola University of Chicago VI-7
SEIGFRIED, Charlene Haddock, Purdue University III-3
SHAPIRO, Michael, Brown University V-1
SHARKEY, Paul, University of Southern Mississippi IX-9
SHERBURNE, Donald, Vanderbilt University I-2
SHRADER-FRECHETTE, Kristin, University of South Florida VI-8
SHUN, Kwong-Loi, University of California/Berkeley IV-2
SIDERITS, Mark, Illinois State University III-1
SILLS, Chip, U S Naval Academy X-2
SILVERMAN, Allan, Ohio State University V-2
SILVERMAN, Hugh, State University of New York/Stony Brook II-1
SIMCO, Nancy, Memphis State University VI-13
SLEIGH, Jr. Robert C., University of Massachusetts/Amherst VI-12
SMITH, John E., Yale University VI-13
SMITH, Kenneth, University of Dallas III-9
SMITH, Patricia G., University of Kentucky V-5

SMITH, Tony, Iowa State University	VIII-5
STERN, David, University of Toledo	III-9
STONE, Robert V , Long Island University/C. W. Post College	XI-2
STURM, Fred, University of New Mexico	IV-4
THOMAS, Laurence, Syracuse University	VII-1
TIMM, Jeffrey, Wheaton College, Massachusetts	I-3
TONG, Rosemarie, Davidson College, North Carolina	V-5, IX-7
TUCKER, John, University of North Florida	VI-11
TYMIENIECKA, Anna-Teresa, World Phenomenology Institute	II-5
VARNER, Gary E., Texas A & M University	IV-3
VER EECKE, Wilfried, Georgetown University	V-8
WALSH, George V., Salisbury State College, Maryland	VIII-1
WALSH, Margaret, Bradford College, Massachusetts	VI-10
WATSON, James, Loyola University	VI-2
WAYMAN, Alex, Columbia University	I-3
WEINSTEIN, Mark, Montclair State University	VI-1
WEISSMAN, David, City University of New York/City College	VI-13
WENZ, Peter S., Sangamon State University	IV-3
WERNER, Richard, Hamilton College, New York	VI-14
WERTHEIMER, Alan, University of Vermont	IV-1
WEST, Robin, University of Maryland/College Park	V-5
WESTPHAL, Merold, Fordham University	III-4
WESTRA, Laura, University of Windsor, Canada	VI-8, IX-5
WHITE, Alan, Williams College, Massachusetts	VIII-6
WIDER, Kathleen, University of Michigan/Dearborn	III-5
WILKINSON, James, University of Kentucky	X-2
WILLIAMS, David, University of Sheffield, England	II-1
WILLIAMS, Joan, American University	V-5
WILLIAMS, Robert R., Hiram College, Ohio	III-4
WILSHIRE, Bruce, Rutgers University	IX-8
WILSON, Catherine, University of Oregon	VI-3
WINFIELD, Richard, D., University of Georgia	VIII-5
WINKLER, Kenneth, Wellesley College, Massachusetts	VI-3
WOODS, John, University of Lethbridge, Canada	VI-1
WYSCHOGRAD, Edith, City University of New York/Queens College	VI-2
YAKE, J. Stanley, State University of New York/Albany	II-8
YAMAMOTO, Yutaka, University of New Hampshire	VI-11
YOUNG, Iris, University of Pittsburg	III-8
ZLOTKOWSKI, Edward, Bentley College, Massachusetts	IX-5
ZOELLER, Guenter, University of Iowa	VI-5

The Journal of Philosophy and Hackett Publishing Company invite you to review these distinguished recipients of the JOHNSONIAN PRIZE IN PHILOSOPHY

CHRISTINE SWANTON

Freedom:

A Coherence Theory

March 1992 224 pp $29.50 cloth
prof price $18.00
awarded the 1990 prize

"The best philosophical work on freedom that I have ever read. Swanton does for the concept of freedom what Rawls did for the concept of justice . . . A most original combination of aspects of Rawls's wide reflective equilibrium, Aristotle's appeal to endoxa, and the Wittgensteinian demand that, in giving an account of a concept, we must consider its point."
— R. Hursthouse, *The Open University*

C. L. HARDIN

Color for Philosophers:

Unweaving the Rainbow

Foreword by Arthur Danto
268 pp $16.00 paper prof price $12.50
awarded the 1986 prize

"Much the best philosophically oriented book about colour that has been written."
— Jonathan Westphal, in *Mind*

ROBERT DUNN

The Possibility of Weakness of Will

192 pp $14.50 paper prof price $9.50
awarded the 1984 prize

"The book reveals a rare blend of virtues: a mastery of the literature, a lucidity of mind, a rigor of argument, and a robust sense of psychological realism."
— Bela Szabados, in
Canadian Philosophical Review

VANN McGEE

Truth, Vagueness, & Paradox:

An Essay on the Logic of Truth

256 pp $30.00 prof price $18.50
awarded the 1988 prize

"McGee's own account of truth is not only a significant technical achievement in logic, but also a philosophical achievement, offering a new way of looking at the putatively universal notion of truth (or truth in our language)."
— Charles Parsons, *Harvard University*

HARRY van der LINDEN

Kantian Ethics & Socialism

384 pp $16.50 paper prof price $12.50
awarded the 1985 prize

"Throws valuable light on the three greatest modern German political thinkers: Kant, Hegel, and Marx."
— Patrick Riley, *University of Wisconsin*

MORELAND PERKINS

Sensing the World

362 pp $16.50 paper prof price $12.50
awarded the 1982 prize

"The most important philosophical work on perception since Price's famous work."
— W. P. Alston, *Syracuse University*

CHARLES GUIGNON

Heidegger & the Problem of Knowledge

269 pp $12.75 paper prof price $9.50
honorable mention 1981

"An admirably clear account of Heidegger's relation to the philosophical tradition, and especially of his criticism of Cartesianism."
— Richard Rorty, *University of Virginia*

◊ Hackett Publishing Co. ◊

Box 44937, Indianapolis, IN 46244-0937
Telephone 317 635-9250 Fax 317 635-9292

◊ NEW AND FORTHCOMING TITLES ◊

Other Selves

Philosophers on Friendship

Edited by MICHAEL PAKALUK

1991 288 pp $7.95 paper $34.95 cloth
prof price $6.00 exam price $3.00

"A valuable, long-needed anthology that serves several purposes well—a supplementary text to courses in ethics, social and political philosophy, philosophy of community and of law."
— Mary Rousseau, *Marquette University*

Includes selections from Plato, Aristotle, Cicero, Seneca, Aelred, Aquinas, Montaigne, Bacon, Kant, Emerson, Kierkegaard, and E. Telfer.

HEGEL

The Encyclopaedia Logic

Part I of the *Encyclopaedia of the Philosophical Sciences* with the *Zusätze*

Translated, with introduction and notes, by T. F. GERAETS, W. A. SUCHTING, & H. S. HARRIS

Oct 1991 416 pp $14.95 paper
$34.95 cloth prof price $12.50

The first new translation of Hegel's fundamental text in more than a century, this is the collaborative effort of three distinguished Hegel scholars who have rendered Hegel's logical terminology with unsurpassed precision and fluency.

Central Themes in Early Modern Philosophy: Essays Presented to Jonathan Bennett

Edited by J. COVER & M. KULSTAD

1990 350 pp $16.95 paper $34.95 cloth
prof price $12.50

"A selection of some of the best work being done in early modern philosophy by Anglo-American philosophers today."
— Daniel Garber, *University of Chicago*

PLATO

Protagoras

Translated by S. LOMBARDO & K. BELL

Introduction by MICHAEL FREDE

March 1992 112 pp.
$4.95 paper (estimated) $27.50 cloth

"Conveys both the philosophical and the dramatic context better than any existing translation. . I am very impressed with the vividness and the easy flow of the prose."
— John Cooper, *Princeton University*

new edition

***Classics of Western Philosophy*, 3d ed.**

Edited by STEVEN M. CAHN

1990 1190 pp. $27.50 paper
$45.00 cloth
prof price $18.50 exam price $5.00

Now with 80 pages of Plato's *Republic*

new edition

CORNMAN, LEHRER, & PAPPAS

***Philosophical Problems & Arguments: An Introduction*, 4th ed.**

March 1992 384 pp $16.95 paper
$32.50 cloth
prof price $10.00 exam price $3.00

new edition

LEIBNIZ

Discourse on Metaphysics and Other Essays

Translated by D. GARBER & R. ARIEW

Nov. 1991 96 pp. $4.95 paper
exam price $1.50

Hackett Publishing Company

Box 44937, Indianapolis, IN 46244-0937

Telephone 317 635-9250 Fax 317 635-9292

◊ AND A MAJOR NEW TEXT ◊

Classics of Moral and Political Theory

Edited by MICHAEL L. MORGAN

March 1992 1280 pp $27.50 paper $48.50 cloth
exam price $5.00 prof. price $18.50

The most comprehensive anthology available for teaching the history of Western moral and political theory.

All texts complete except where noted

◊

PLATO *Apology, Crito, Republic* tr. by Grube

ARISTOTLE *Nicomachean Ethics* tr. by Irwin; *Politics* (Bk I, 1–9, Bk II, 1–5, 9, Bk III, 1–2, 4–5; Bk VII, 1–3, 13–15) tr. by Jowett, rev. by Barnes

EPICURUS *Letter to Menoeceus, Principal Doctrines* tr. by Inwood & Gerson

EPICTETUS *Encheiridion* tr. by N. P. White

AUGUSTINE *City of God* (Bk XIX, 1, 4–8, 14,17, 19, 27) tr. pub. by Fathers of the Church

AQUINAS *On Kingship, On Law* (from the *Summa Theologica*) tr. by Fathers of the English Dominican Province, rev. by Baumgarth & Regan

MACHIAVELLI *The Prince, Discourses* (Bk I, 1–2, Bk II, 1–2, 20, 29; Bk III, 1, 9), Letter to Vettori tr. by Bondanella & Musa

HOBBES *Leviathan* (Introduction, Pt I–II, Conclusion)

LOCKE *Second Treatise of Government*

HUME *Treatise of Human Nature* (Bk III, Pt I, i–ii, Pt II, i–ii)

ROUSSEAU *Discourse on the Origin of Inequality, On the Social Contract* tr. by Cress

KANT *Grounding for the Metaphysics of Morals* tr. by Ellington

MILL *On Liberty, Utilitarianism*

MARX *Alienated Labor, On the Jewish Question; Communist Manifesto, Critique of the Gotha Program* tr. pub. by International Publishers

NIETZSCHE *On the Genealogy of Morals* (Preface, 1st and 2nd Essays) tr. by Kaufmann

Visit us at APA Eastern booths 402, 404, & 406.

Hackett Publishing Company

PHILOSOPHY PAST AND PRESENT
NEW from PENGUIN USA

LUDWIG WITTGENSTEIN
The Duty of Genius
Ray Monk

A unique insight into the life and work of Wittgenstein, one of the most acute philosophical minds of the twentieth century. "Well written, intelligent and moving...a first-rate biography that stands as both a fine piece of scholarship and a vivid portrait."—*The New York Times.* 672 pp.

Penguin paperback
0-14-015995-9 $15.95
Available November 1991

A NEW VIEW OF SOCIETY AND OTHER WRITINGS
Robert Owen

Edited and Introduced by Gregory Claeys. A diverse selection of inspiring writings from Robert Owen, one of the greatest British social reformers, reflecting his intense concern for equality and justice through his radical views of British society, working practices, education, and the Poor Laws. 416 pp.

Penguin Classic paperback
0-14-043348-1 $9.95

PERSONALITY
An Introduction to The Theories of Psychology
Peter Morea

In this accessible volume, Morea examines and explains the main theories of personality through a survey of the well-established approaches of Freud, Jung, Skinner, and others, providing a comprehensive survey of their insights and ideas and an objective and multi-sided understanding of the subject. 192 pp.

Penguin paperback
0-14-012519-1 $7.95

THE ENNEADS
Plotinus

Translated by Stephen MacKenna. Edited by John Dillon. This interpretative translation of Plotinus' highly original synthesis of Platonism, mystic passion, Greek philosophy, and variants of the Trinity and other central doctrines offers profound insights into an often inaccessible yet brilliant thinker. 688 pp.

Penguin Classic paperback
0-14-044520-X $12.95
Available November 1991

THE FIRE AND THE SUN
Why Plato Banished the Artists
Iris Murdoch

Based on her 1976 Romanes Lecture at Oxford University, Murdoch's latest is a succinct, witty book which discusses Plato's attitude toward art and the age-old quarrel between poetry and philosophy. Providing an accessible and necessary compression of Plato's thought, *The Fire and the Sun* also includes observations on Kant, Tolstoy, Freud, and Kierkegaard. 96 pp.

Viking hardcover
0-670-83708-3 $14.95

JOURNEY THROUGH GENIUS
The Great Theorems of Mathematics
William Dunham

This rare combination of the historical, biographical, mathematical, and philosophical offers a captivating introduction to a neglected, important field of human creativity. "A fascinating lingering over individual examples of ingenuity and insight."—Isaac Asimov. 448 pp.

Penguin paperback
0-14-014739-X $10.95

Visit us at Booth #201
PENGUIN USA
Academic Marketing Department, 375 Hudson Street, New York, NY 10014-3657

IF IT SPARKS DEBATE, IT'S FROM

PROMETHEUS

Prescription: Medicide

The Goodness of Planned Death

Jack Kevorkian, M.D.

Kevorkian made headlines in 1990 when he aided a victim of rapidly degenerating Alzheimer's disease, in performing the first publicly acknowledged physician-assisted suicide—what he terms medicide. He is a man to whom the highest ethical principle is individual self-determination. In this book, he takes on the medical establishment, politicians, and all those who actively resist a rational and comprehensive program of dignified, humane, and beneficial planned death. **245 pages • Cloth $23.95**

The Gun Control Debate

You Decide

edited by Lee Nisbet

"An excellent point of departure for students beginning to study the gun control issue."

—**Library Journal**

341 pages • Paper $16.95

The Ethics of Abortion

Pro Choice vs. Pro Life

edited by Robert M. Baird and Stuart E. Rosenbaum

"A timely recap of the issues surrounding abortion . . . provides much food for thought and discussion on a vital current affairs topic." —**Booklist**

172 pages • Paper $14.95

Euthanasia

The Moral Issues

edited by Robert M. Baird and Stuart E. Rosenbaum

"A significant contribution toward defining euthanasia, and describing and distinguishing among its various forms . . . challenging."

—**Booklist**

182 pages • Paper $11.95

Animal Experimentation

The Moral Issues

edited by Robert M. Baird and Stuart E. Rosenbaum

"Well written and documented."

—**Library Journal**

182 pages • Paper $14.95

Order Toll Free (800) 421-0351 (24 Hours)

or write to

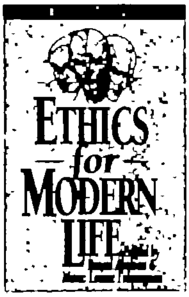
Prometheus Books

59 John Glenn Drive • Amherst, NY 14228-2197

Add $3.00 p&h for first book, $1.50 for each additional book ($9.00 maximum).

NY state residents add sales tax. #40121

ST. MARTIN'S PRESS



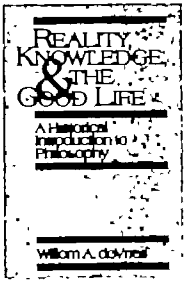
ETHICS FOR MODERN LIFE

Fourth Edition

RAZIEL ABELSON and MARIE-LOUISE FRIQUEGNON

A well balanced introduction to both theoretical and applied ethics that first examines the standards of moral right and wrong and then applies these basic philosophic principles to modern social issues. The chapter on environmental ethics is new to this edition, and the chapter on suicide has been reinstated from the second edition.

Paperbound. 480 pp. 1991



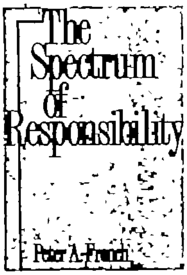
REALITY, KNOWLEDGE, AND THE GOOD LIFE

A Historical Introduction to Philosophy

WILLEM A. DEVRIES

A text-reader containing 31 selections from 18 major philosophers. The text presents three major themes in the history of philosophy – reality, knowledge, and the good life – and shows how the themes have evolved over time in the writings of important philosophers from the early Greeks to the 20th century.

Paperbound. 704 pp. 1991



THE SPECTRUM OF RESPONSIBILITY

PETER A. FRENCH

A brief anthology that brings together differing viewpoints on the concept of moral responsibility – specifically, the responsibility of individuals and groups to each other and to society as a whole. Responsibility is examined at both the collective and personal level, offering students the opportunity to see moral roles in the business or professional world.

Paperbound. 236 pp. 1991

SMP ST. MARTIN'S PRESS
College Division, Department JR
175 Fifth Avenue, New York, N.Y. 10010
College Desk 1-800-446-8923

PHILOSOPHY
FROM FORDHAM

Visit us at APA Booth #109 to see our full offering of philosophy titles

---

Practical Knowledge

Yves Simon

Simon, one of this century's greatest students of the virtue of practical wisdom, explores a range of subjects from ultimate theoretical and foundational concerns to the most concrete and immediate questions regarding the role of practical wisdom in personal and social decision-making.

176 pages $35.00 cloth $19.95 paper

Soundings in St. Augustine's Imagination

Robert J. O'Connell, S.J.

A leading Augustinian scholar explores the powerful poetic imagery which infuses St. Augustine's philosophical and theological works.

296 pages $40.00 cloth $19.95 paper

Modernity and Its Discontents

John D. Caputo, James L. Marsh and Merold Westphal

In point/counter-point format, this volume presents the debate between Caputo's and Marsh's philosophical positions of modernism and post-modernism. The timely debate opens with each debater's critique of the other's work. At the core of the argument lies the question of the role each person can play in creating a truly humane society.

240 pages $32.50 cloth $19.95 paper

Hermeneutics and Human Finitude

Toward a Theory of Understanding

P. Christopher Smith

Smith elaborates on Gadamer's work in seeking a recollection of shared ethical principles, which alone may make possible some measure of consensus.

296 pages $40.00 cloth

A Reading of Hegel's *Phenomenology of Spirit* Revised Edition

Quentin Lauer, S.J.

From the First Edition:

"It is by far the best interpretation and exposition of Hegel's book that has ever come my way."

Philosophical Books

303 pages $30.00 cloth $17.95 paper

Frontiers of Consciousness

Foundational Studies in Classical American Philosophy and Poetry

Stanley J. Scott

Scott's work is a true interdisciplinary study-- of the American philosophers who have been called the Builders and the work of three great modernist poets.

156 pages $40.00 cloth

To Order: Call: 1-800-666-2211 Fax: 1-800-688-2877

To receive a catalogue please write to:

FORDHAM
UNIVERSITY PRESS
University Box L, Bronx, NY 10458-1572 (212) 579-2319

CORNELL

Perceiving God
The Epistemology of Religious Experience
By **WILLIAM P. ALSTON.** $34.50

What Can She Know?
Feminist Theory and the Construction of Knowledge
By **LORRAINE CODE.** $42.50 cloth, $14.95 paper

Identity\Difference
Democratic Negotiations of Political Paradox
By **WILLIAM E. CONNOLLY.** $27.50

Definitions of Art
By **STEPHEN DAVIES.** $31.50 cloth, $12.95 paper

Configurations of Masculinity
A Feminist Perspective on Modern Political Theory
By **CHRISTINE DI STEFANO**
$29 95 cloth, $10 95 paper

Cornell Studies in the Philosophy of Religion—
Time and Eternity
By **BRIAN LEFTOW.** $42.95

The Liberal Self
John Stuart Mill's Moral and Political Philosophy
By **WENDY DONNER.** $35.00 cloth, $12.95 paper

Milton among the Philosophers
Poetry and Materialism in Seventeenth-Century England
By **STEPHEN M. FALLON** $34.50

Racism and Justice
The Case for Affirmative Action
By **GERTRUDE EZORSKY.** $21.95 cloth, $6.95 paper

A Peculiar Fate
Metaphysics and World-History in Kant
By **PETER D. FENVES.** $36.50

Socrates, Ironist and Moral Philosopher
By **GREGORY VLASTOS.** *Townsend Lectures Series*
Cornell Studies in Classical Philology.
$57.50 cloth, $16.95 paper

Politics/Sense/Experience
A Pragmatic Inquiry into the Promise of Democracy
By **TIMOTHY V. KAUFMAN-OSBORN.** $29.95

Music Alone
Philosophical Reflections on the Purely Musical Experience
By **PETER KIVY.** *New in Paper!*
$10.95

Sound and Semblance
Reflections on Musical Representation
By **PETER KIVY.** *New in Paper!*
$12 95

Vico's Science of Imagination
By **DONALD PHILLIP VERENE.**
New in Paper! $10.95

BOOTH 417

Postmodernism and Its Critics

By **JOHN MCGOWAN.** $39.95 cloth, $12.95 paper

Ancient Commentators on Aristotle—

On Aristotle's *Categories*

By **AMMONIUS.** Translated by **S. Marc Cohen** and **Gareth B. Matthews.** $46.50

On Aristotle's *Prior Analytics 1.1–7*

By **ALEXANDER OF APHRODISIAS.** Translated by **Jonathan Barnes, Susanne Bobzien, Kevin Flannery,** and **Katerina Ierodiakonou.** $47.95

Place, Void, and Eternity

By **PHILOPONUS** and **SIMPLICIUS.** Translated by **David Furley** and **Christian Wildberg.** $44.50

On Aristotle on the Intellect (*De Anima 3.4—8*)

By **PHILOPONUS.** Translated by **William Charlton.** $52.50

Thomas Reid on Freedom and Morality

By **WILLIAM L. ROWE.** $26.95

Heidegger's Pragmatism

Understanding, Being, and the Critique of Metaphysics

By **MARK OKRENT.** *New in Paper!* $12.95

Whose Science? Whose Knowledge?

Thinking from Women's Lives

By **SANDRA HARDING.** $34.50 cloth, $12.95 paper

John Dewey and American Democracy

By **ROBERT B. WESTBROOK.** $29.95

Herder's Aesthetics and the European Enlightenment

By **ROBERT E. NORTON.** $35.95

The Interpretive Turn

Philosophy, Science, Culture

Edited by **DAVID R. HILEY, JAMES BOHMAN,** and **RICHARD SHUSTERMAN.** $39 95 cloth, $12.95 paper

Primary *Ousia*

An Essay on Aristotle's Metaphysics *Z and H*

By **MICHAEL J. LOUX.** $39 95

Unreality

The Metaphysics of Fictional Objects

By **CHARLES CRITTENDEN.** $26.50 cloth, $9.95 paper

CORNELL
University Press

124 Roberts Place, Ithaca NY 14850

NEW FROM PENN STATE PRESS

MISSING SOCRATES
Problems of Plato's Writing
JAY FARNESS
216 pages **$29.50**

PLATO'S *PARMENIDES*
The Conversion of the Soul
MITCHELL H. MILLER, JR.
314 pages **paper: $14.95**

SELECTED WORKS, 1764–1767
Early Addresses, Essays, and Drafts; Fragments on Recent German Literature
JOHANN GOTTFRIED HERDER
Edited by Ernest A. Menze and Karl Menges
Translated by Ernest A. Menze with Michael Palma
368 pages **$35.00**

REORIENTING RHETORIC
The Dialectic of List and Story
JOHN D. O'BANION
312 pages **$35.00**
Forthcoming for Spring 1992

ENGENDERING DEMOCRACY
ANNE PHILLIPS
200 pages
cloth: $28.50 paper: $13.95

FEMINISM CONFRONTS TECHNOLOGY
JUDY WAJCMAN
220 pages
cloth: $28.50 paper: $13.95

KIERKEGAARD'S CRITIQUE OF REASON AND SOCIETY
MEROLD WESTPHAL
142 pages **paper: $12.95**
Forthcoming for Spring 1992

MALLARMÉ, OR THE POET OF NOTHINGNESS
JEAN-PAUL SARTRE
Translated and Introduced by Ernest Sturm
150 pages **paper: $12.95**

SEE THESE BOOKS
ON DISPLAY
IN THE AAUP BOOTH

PENN STATE PRESS
Suite C, 820 N. University Drive
University Park, PA 16802
(814)865-1327

Princeton

New in paperback

Aristotle on the Human Good

Richard Kraut

Aristotle's *Nicomachean Ethics*, which equates the ultimate end of human life with happiness (*eudaimonia*), is thought by many readers to argue that this highest goal consists in the largest possible aggregate of intrinsic goods. Richard Kraut proposes instead that Aristotle identifies happiness with only one type of good: excellent activity of the rational soul.

"*Aristotle on the Human Good* will be the starting-point of all subsequent discussion of the subject, especially because of the fullness and detail of its arguments."
—*Christopher Rowe, The Times Literary Supplement*
Now in paper: $16.95 ISBN 0-691-02071-X

New in paperback

Aristotle on Substance

The Paradox of Unity

Mary Louise Gill

This book explores a fundamental tension in Aristotle's metaphysics: how can an entity such as a living organism—a composite generated through the imposition of form on preexisting matter—have the conceptual unity that Aristotle demands of primary substances? Mary Louise Gill bases her treatment of the problem of unity, and of Aristotle's solution, on a fresh interpretation of the relation between matter and form.

"A substantial and philosophically sophisticated contribution to the growing literature on Aristotelian metaphysics. . . . Focusing on the role of matter and the persistence of matter in explanations of the generation and constitution of compounds, Gill argues that Aristotle offers a unified and metaphysically defensible account of the identity and continuity of generated substances. Her argument is carried on at a very high level throughout and reflects a thorough absorption of recent scholarship."—*Choice*
Now in paper: $14.95 ISBN 0-691-02070-1

Princeton University Press

41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 258-4900
ORDERS: 800-PRS-ISBN (777-4726) • OR FROM YOUR LOCAL BOOKSTORE

Westview

◆ **After the Demise of the Tradition**
Rorty, Critical Theory, and the Fate of Philosophy
Kai Nielsen
April 1991 • 278 pp. • $38.50 hc

◆ **Pragmatism in Law and Society**
edited by Michael Brint and William Weaver
November 1991 • ca. 224 pp. • $49.95 hc • $15.95 pb

◆ **Virtues and Rights**
The Moral Philosophy of Thomas Hobbes
R.E. Ewin
September 1991 • 223 pp. • $38.50 hc

◆ **Morality: What's in it for Me?**
A Historical Introduction to Ethics
William N. Nelson
December 1990 • 166 pp. • $36.95 hc • $10.95 pb

◆ **Fair Play**
Sports, Values, and Society
Robert L. Simon
November 1991 • ca. 210 pp. • $37 hc • $14.95 pb

◆ **Sexual Democracy**
Women, Oppression, and Revolution
Ann Ferguson
May 1991 • 293 pp. • $42.50 hc • $15.95 pb

◆ **Secession**
The Morality of Political Divorce from Fort Sumter to Lithuania and Quebec
Allen Buchanan
September 1991 • ca. 168 pp. • $34.95 • $14.95 pb

◆ **Feminist Legal Theory**
Readings in Law and Gender
edited by Katharine T. Bartlett and Rosanne Kennedy
November 1991 • ca. 383 pp. • $55 hc • $16.95 pb

◆ **Wittgenstein and Legal Theory**
edited by Dennis M. Patterson
January 1992 • ca. 256 pp. • $49.95 hc

◆ **Revolutions in Knowledge**
Feminism in the Social Sciences
edited by Sue Rosenberg Zalk and Janice Gordon-Kelter
November 1991 • ca. 224 pp. • $44 hc

◆ **The Opened Curtain**
A U.S.-Soviet Philosophy Summit
edited by Keith Lehrer and Ernest Sosa
July 1991 • 318 pp. • $38.50 sc

Westview Press
5500 Central Avenue • Boulder, CO 80301
Phone: (303) 444-3541 • FAX: (303) 449-3356

KLUWER ACADEMIC PUBLISHERS

Philosophy Publishing Program

Publishing in the areas of:

Philosophy of Science	Epistemology
Logic	Philosophy of Medicine
Phenomenology	Philosophy of Law
History of Philosophy	*etc.*

With such major series as:

Synthese Library
edited by Jaakko Hintikka
Boston Studies in the Philosophy of Science
edited by Robert S. Cohen
Philosophy and Medicine
edited by H. Tristram Engelhardt, Jr. and Stuart Spicker
Clinical Medical Ethics
edited by H. Tristram Engelhardt, Jr.
Philosophy and Technology
edited by Paul Durbin
International Archives of the History of Ideas
edited by P. Dibon and R. Popkin
Analecta Husserliana
edited by Anna-Teresa Tymieniecka
Contributions to Phenomenology
edited by William R. McKenna
Husserliana Gesammtausgabe
and Edmund Husserl Collected Works
Western Ontario Series
edited by Robert E. Butts
Phaenomenologica
edited by the Husserl Archives at Louvain
Nijhoff International Philosophy Series
edited by Jan Srzednicki
Philosophical Studies Series
edited by Keith Lehrer
The New Synthese Historical Library
edited by Norman Kretzmann

Contact us for more information!

Visit our booth at APA Eastern - #'s 208 & 210

Kluwer Academic Publishers · 101 Philip Drive · Norwell, MA 02061
Phone 617-871-6600 · Fax 617-871-6528 · E-mail: Kluwer@world.std.com.

NEW FROM BLACKWELL

HOLISM

A Shopper's Guide

JERRY FODOR & ERNEST LEPORE

This book sets out the main arguments for semantic holism and subjects them to detailed critical evaluation. The implications of holism concerning meaning for other philosophical issues are also explored.

December	*156 pages*	
0-631-18193-8	*paper*	*$14.95*
0-631-18192-X	*cloth*	*$36.95*

THE IRIGARAY READER

Edited by MARGARET WHITFORD

Translated by David Macey

This collection of Luce Irigaray's most important papers ranges across feminism, philosophy, psychoanalysis and linguistics. With a number of pieces appearing for the first time in English, this is the most comprehensive collection of her work.

November	*252 pages*	
0-631-17043-X	*paper*	*$19.95*
0-631-17042-1	*cloth*	*$39.95*

HIPPARCHIA'S CHOICE

An Essay Concerning Women and Philosophy

MICHELE LE DOEUFF

Enormously well-received and widely reviewed upon its publication in France, *Hipparchia's Choice* is a work of rare irreverence and unimpeachable seriousness. It points the way to a philosophy accountable to history, feminism, society and intelligibility.

October	*368 pages*	
0-631-17641-1	*paper*	*$19.95*
0-631-17639-X	*cloth*	*$49.95*

THE PRODUCTION OF SPACE

HENRI LEFEBVRE

Translated by Donald Nicholson-Smith

Afterword by David Harvey

"The Production of Space ***is a pathbreaking text which might well play a decisive role in reorienting the problems and debates of contemporary philosophy."***

— F.R. Jameson, Duke University

August	*448 pages*	
0-631-18177-6	*paper*	*$19.95*
0-631-14048-4	*cloth*	*$54.95*

MODERNITY AND IDENTITY

Edited by SCOTT LASH & JONATHAN FRIEDMAN

Containing important new essays by Richard Rorty, Marshall Berman and Martin Jay, this groundbreaking work announces a radical new departure within contemporary debates on modernism and postmodernism.

November	*448 pages*	
0-631-17586-5	*paper*	*$19.95*
0-631-17585-7	*cloth*	*$64.95*

THE TRUTH ABOUT RELATIVISM

JOSEPH MARGOLIS

The first full-length defense of philosophical relativism, this book assembles with great care the principal lines of attack in the ancient world and the various lines of attack that have been developed in contemporary Anglo-American and continental European philosophy.

November	*240 pages*	
0-631-18178-4	*paper*	*$19.95*
0-631-17911-9	*cloth*	*$44.95*

BLACKWELL
Publishers

Three Cambridge Center, Cambridge, MA 02142
Call Toll-Free (800) 445-6638
Vermont & Canada dial (802) 878-0315 (MasterCard/Visa only)

NEW FROM BLACKWELL

Philosophy and the Idea of Freedom

ROY BHASKAR

In the first volume of a major new three volume study on the trajectory of the western philosophical tradition and the state of contemporary philosophy, the author sets out to develop a critique of the work of Richard Rorty and shows how Rorty falls victim to the epistemological problematic he himself describes.

October	*200 pages*	
0-631-17082-0	*paper*	*$19.95*
0-631-15911-8	*cloth*	*$47.95*

The Analytic Ambition

An Introduction to Philosophy

WILLIAM CHARLTON

In discussing several central philosophical problems, William Charlton shows why they are problems for philosophy rather than for any other discipline and thereby illustrates and confirms a general theory of the nature and scope of philosophical enquiry.

November	*256 pages*	
0-631-16935-0	*paper*	*$19.95*
0-631-16934-2	*cloth*	*$47.95*

Realism and Truth

Second Edition

MICHAEL DEVITT

The definition of "realism" in the external world is a question that is commonly conflated with the semantic question of truth. This book distinguishes these questions and argues, from a naturalist perspective, that the metaphysical one is anterior.

December	*288 pages*	
0-631-17551-2	*paper*	*$27.95*

Philosophy

A Beginner's Guide

JENNY TEICHMAN &
KATHERINE C. EVANS

"Compact yet thorough...it covers more topics, in a short but effective manner, than any other general introduction I have ever seen."

—George Graham, University of Alabama at Birmingham

July	*288 pages*	
0-631-17131-2	*paper*	*$18.95*
0-631-17129-0	*cloth*	*$44.95*

BLACKWELL JOURNALS

From 1992 Blackwell Publishers will be publishing *NOUS* and *The Heythrop Journal.*

NOUS offers a stimulating exploration of philosophic problems. Published five times a year, it presents papers and essays regardless of the authors philosophic school or point of view.

The Heythrop Journal provides a forum for scholars in the fields of philosophy and theology, and encourages interdisciplinary comment and debate.

Please visit booths 103 and 105 at the APA for more information on these and other Blackwell Journals.

APA MEMBERS RECEIVE A 15% DISCOUNT ON ALL JOURNAL SUBSCRIPTIONS

BLACKWELL
Publishers

Three Cambridge Center, Cambridge, MA 02142
Call Toll-Free (800) 445-6638
Vermont & Canada dial (802) [illegible] (MasterCard/Visa only)

Philosophy of ROUTLEDGE

CREATIVE CHARACTERS

Elisabeth Young-Bruehl

In *Creative Characters*, Young-Bruehl argues that creative people develop an image of their characters which they both aspire to and project into whatever they create. This character-ideal appears in their works, their social visions, their philosophies of nature, and also their understandings of creative processes, their own and others'.

272 pp $25.00/cloth

ART, MIMESIS AND THE AVANT GARDE

Aspects of a Philosophy of Difference

Andrew Benjamin

Art, Mimesis and the Avant Garde explores the relationship between art and philosophy. Benjamin argues for a reworking of the task of philosophy in terms of the centrality of ontology. It is in relation to this centrality— understood through the differences between modes of being—that art, mimesis, and the avant-garde come to be presented

7 b/w photographs
256 pp $49.95/cloth $18.95/paper

FIRST PERSON PLURAL

Multiple Personality and the Philosophy of Mind

Stephen E. Braude

First Person Plural is the first full-length philosophical examination of multiple personality disorder (MPD) and its significance for a variety of classic issues in the philosophy of mind.

288 pp $49.95/cloth

HÉLÈNE CIXOUS

A Politics of Writing

Morag Shiach

In *Hélène Cixous*, Shiach provides an introduction to the range of Cixous' creative and theoretical writings, clarifying the philosophical and historical context of her work.

6 b/w photographs
224 pp $49.95/cloth $13.95/paper

HEGEL AND LEGAL THEORY

Edited by **Drucilla Cornell, Michel Rosenfeld,** and **David Gray Carlson**

336 pp $45.00/cloth $15.95/paper

The Arguments of the Philosophers series

Now in paper!

DEWEY

J. E. Tiles

284 pp $16.95/paper

Now in paper!

AUGUSTINE

Christopher Kirwan

256 pp $16.95/paper

Now in paper!

J. L. AUSTIN

G. J. Warnock

174 pp $14.95/paper

LOCKE

Vol 1: Epistemology
Vol 2: Ontology

Michael Ayers

Two Volume Set
704 pp $127.50/cloth

FREEDOM, TRUTH AND HISTORY

An Introduction to Hegel's Philosophy

Stephen Houlgate

Houlgate offers a comprehensive introduction to the philosopher G. W. F. Hegel, now recognized as one of the most important modern thinkers.

3 b/w photographs
288 pp $49.95/cloth $16.95/paper

TRUTH AND EROS

Foucault, Lacan and the Question of Ethics

John Rajchman

In *Truth and Eros*, Rajchman isolates the question of ethics in the work of Foucault and Lacan and explores its ramifications and implications for the present day.

192 pp $42.50/cloth $13.95/paper

PATTERNS OF DISSONANCE

A Study of Women in Contemporary Philosophy

Rosi Braidotti

Translated by **Elizabeth Guild**

"Braidotti's tour de force shows the immense theoretical impact of recent French and American feminist interventions in Western philosophy .... This important study will transform the ways in which both pre-feminist and feminist thinkers evaluate feminist theory, Western philosophy, and, especially, feminist philosophy."
—Sandra Harding,
University of Delaware

324 pp $49.95/cloth $16.95/paper

FRENCH PHILOSOPHERS IN CONVERSATION

Levinas, Schneider, Serres, Irigaray, Le Doeuff, Derrida

Raoul Mortley

"... each [interview] is spirited, interesting, and accessible to the educated general reader."
—*Choice*

108 pp $42.50/cloth $12.95/paper

DEATH AND DESIRE

Psychoanalytic Theory in Lacan's Return to Freud

Richard Boothby

Death and Desire presents in Lacanian terms a new integration of psychoanalytic theory in which the battery of key Freudian concepts—from the dynamics of the Oedipus complex to the topography of ego, id and superego—are seen to intersect in Freud's most far-reaching and speculative formulation of a drive towards death.

256 pp $42.50/cloth $15.95/paper

The Problems of Philosophy: Their Past and Present

Now in paper!

PRACTICAL REASONING

Robert Audi

256 pp $15.95/paper

KNOWLEDGE OF THE EXTERNAL WORLD

Bruce Aune

259 pp $39 95/cloth

Now in paper!

MORAL KNOWLEDGE

Alan H. Goldman

240 pp $14.95/paper

Now in paper!

PERSONAL IDENTITY

Harold W. Noonan

304 pp $15.95/paper

WRITING AND THE MORAL SELF

Berel Lang

"Berel Lang is a thoughtful and thought-provoking writer— civilized, lucid, and learned. He writes about important subjects, and he stimulates one to a pleased agreement and useful dissent." —Irving Howe

160 pp $42.50/cloth $13.95/paper

Now in paper!

WITTGENSTEIN AND MORAL PHILOSOPHY

Paul Johnston

272 pp $15.95/paper

GADAMER AND HERMENEUTICS

Edited and with an introduction by

Hugh J. Silverman

Gadamer and Hermeneutics provides a series of essays debating the virtues of Gadamerian hermeneutics versus aspects of the work of Heidegger, Ricoeur, Barthes, Derrida and Habermas, with each essay shedding special light on the hermeneutic process.

288 pp $45 00/cloth $15.95/paper

Thinking Gender

BEYOND ACCOMMODATION

Ethical Feminism, Deconstruction, and the Law

Drucilla Cornell

In ***Beyond Accommodation***, Cornell offers a highly original vision of what feminist theory can offer contemporary women. A legal scholar whose work has roots in European philosophy, Cornell shows how a way to ethical feminism can be found through an alliance with deconstruction.

240 pp $45.00/cloth $14.95/paper

DISCIPLINING FOUCAULT

Feminism, Power and the Body

Jana Sawicki

In ***Disciplining Foucault***, Sawicki argues that a Foucauldian feminism is possible, rejecting the view that power of phallocentric discourse is total. Instead, like Foucault, she sees discourse as ambiguous and plurivocal, a site of conflict and contestation.

144 pp $42.50/cloth $13.95/paper

MICHEL FOUCAULT, PHILOSOPHER

Edited and translated by **T. J. Armstrong**

Preface by **Lawrence D. Kritzman**

Michel Foucault, Philosopher brings together original essays by an outstanding group of international scholars who present an array of critical readings of Foucault's work and its impact on western thought.

368 pp $49 95/paper $16.95/paper

Critics of the Twentieth Century series

JÜRGEN HABERMAS

Critic in the Public Sphere

Robert C. Holub

224 pp $45 00/cloth $14.95/paper

LUCE IRIGARAY

Philosophy in the Feminine

Margaret Whitford

"Thanks to Whitford's clear and concise portrayal of Irigaray's thought, serious argument about its full worth and validity can now begin. Whitford has put us all in her debt." —Jay Bernstein

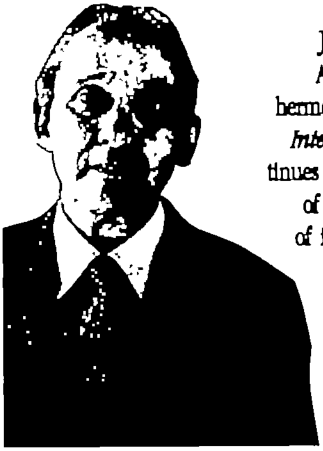
304 pp $45.00/cloth $14 95/paper

Come see us at our booth!

ROUTLEDGE
An imprint of Routledge, Chapman & Hall
29 West 35th Street
New York, NY 10001-2291

Northwestern

PAUL RICOEUR
From Text to Action



Kathleen Blamey and
John B. Thompson, trans.

An essential companion to the hermeneutic classic *The Conflict of Interpretations*, this volume continues and extends Ricoeur's project of constructing a general theory of interpretation. Ricoeur enacts "a gradual reinscription of the theory of texts within the theory of action," moving from a discussion of the importance of critical distanciation to the development of a "hermeneutics of action" and the use of his textual theories to reconsider the relation of ethics to politics. Paper $17.95

PIERRE KLOSSOWSKI
Sade My Neighbor

Alphonso Lingis, trans.

One of the first works to question the philosophical validity of the Marquis de Sade's reason, this seminal work inspired virtually all subsequent study of Sadean thought, including that of de Beauvoir, Deleuze, Derrida, Paulhan, and Lacan. First English translation.
Paper $9.95

Booth #119

NORTHWESTERN UNIVERSITY PRESS
625 Colfax Street Evanston, Illinois 60201 (708) 491-5313

Harcourt Brace Jovanovich College Publishers' Outstanding Selections for 1992

TWENTY QUESTIONS
An Introduction to Philosophy
Second Edition
G. Lee Bowie, *Mount Holyoke College*
Robert C. Solomon, *University of Texas*
Meredith W. Michaels, *Hampshire College*
Paperbound

This popular anthology for introductory courses features a large collection of contemporary selections, including literary and journalistic writing along with established classics. Roughly twenty percent of the readings are new in this edition. A new chapter (12) examines the question "What Does My Race Have to Do with Me?"

THIRTEEN QUESTIONS IN ETHICS
G. Lee Bowie, *Mount Holyoke College*
Kathleen Higgins, *University of Texas*
Meredith W. Michaels, *Hampshire College*
Paperbound

Pressing contemporary ethical questions are explored in readings ranging from journalistic to scholarly in perspective and style. Strong coverage of feminist issues distinguishes this anthology from others.

TO BELIEVE OR NOT TO BELIEVE
Readings in the Philosophy of Religion
E. D. Klemke, *Iowa State University*
Paperbound

Through fifty-six thought-provoking pro and con essays, this innovative textbook explores the essential question of whether religious belief is a viable option for our time.

VALUES AND PUBLIC POLICY
Claudia Mills, *University of Maryland–Baltimore*
Paperbound

This introduction to ethics is composed of articles derived from reports of the Institute for Philosophy and Public Policy. The collection covers a broad range of public policy issues and probes the underlying values that make each issue controversial and important.

HBJ **Harcourt Brace Jovanovich College Publishers**
7555 Caldwell Avenue, Chicago, IL 60648
1-708-647-8822

Announcing a new Oxford journal for 1992

COMMON KNOWLEDGE

Editor
Jeffrey M. Perl

Senior Editor/Fiction Editor
Robert S. Nelsen

Book Review Editor
Beverly Haviland

Poetry Editor
Belle Randall

Assistant Editor
William Kolbrener

Managing Editor
Brooke Allen

Editorial Board
K. A. Appiah
Sissela Bok
Caroline Walker Bynum
Cornelius Castoriadis
Stanley Cavell
William M. Chace
Robert W. Corrigan
Natalie Zemon Davis
Mikhail Epstein
Fang Lizhi
Paul Feyerabend
Elizabeth Fox-Genovese, *Chair*
Joseph Frank
Manfred Frank
Clifford Geertz
Eugene Genovese
Ian Hacking
Pierre Hadot
Francis Haskell
Vicki Hearne
Charles Johnson
Stanley N. Katz
Hugh Kenner
Frank Kermode
Anna Kisselgoff
Leszek Kolakowski
György Konrád
Julia Kristeva
Jacques Le Goff
Greil Marcus
Steven Marcus
Adam Michnik
Czeslaw Milosz
J. G. A. Pocock
Richard Rorty
Abdus Salam
Amartya Sen
Quentin Skinner
Susan Sontag
Jean Stein
G. Thomas Tanselle
Charles Taylor
Tatyana Tolstaya
Stephen Toulmin
Gianni Vattimo

The skeptics, neopragmatists, poststructuralists, feminists, Marxists, historicists, Freudians, paleoconservatives, dissidents, and mavericks who have joined to establish **Common Knowledge** hold that fratricidal partisanship is a legacy of a time that is passing. **Common Knowledge** will publish work in cultural studies and intellectual history that redefines divisive terms and figures of the past and present in ways that make expanded sympathies credible. Now that the Berlin Wall is down, it is time to abandon simplistic distinctions – East versus West, Left versus Right, theory versus praxis, modern versus postmodern. It is time to elaborate a basis for practical, common knowledge.

CK represents a commitment to go beyond established disciplines and, more importantly, beyond the factions at war in the literary and scholarly world. The idea that intellectual life is composed of "issues" and that issues are shaped like battlefields or game boards – with "sides" that one must "take" – is at best obsolete and impractical. CK seeks to refine a new skepticism – a postmodern, post-war doubt of claims to knowledge and authority – that presents no obstacle to solidarity.

CK will publish material that focuses on the timely and the particular rather than on the timeless and the theoretical – material that renders common knowledge possible. The editors solicit articles that reassociate intellectual communities divided by the cold war, articles that establish links among disparate intellectual schools, articles that deepen the conversation within each school, articles that break down the distinction between scholarly and narrative form, and articles that reopen on higher ground debates about incommensurability, ambivalence, indeterminacy, untranslatability, and the will-to-power.

Manuscripts, proposals for articles, and books proposed for review should be sent to **Common Knowledge**, School of Arts and Humanities, The University of Texas at Dallas, Box 830688 – JO 31, Richardson, TX 75083-0688

1992 subscription to Common Knowledge (3 issues) at the special charter rate

[ ] Individuals $19 (reg. $24) [ ] Institutions $39 (reg. $48) Add $14 for subscriptions outside the US

Send order with payment to the Journals Department, Oxford University Press, 2001 Evans Road, Cary, NC 27513. For credit card orders (MasterCard, VISA, and American Express accepted), include account number, expiration date, and signature.

The Journal of Philosophy

Missing Issues?

Complete volumes and all separate issues are available back to Vol. I, No. 1 (January 7, 1904). Prices, not including shipping charges, are as follows:

Single issues:

1904–1976	$3.00
1977–present	$4.00
Dewey Lectures	$5.00

Double issues, devoted to the symposia of the Annual Meetings of the American Philosophical Association, Eastern Division:

1943–1976	$7.00 each set
1977–present	$9.00 each set

Complete volumes:

1904–last year	$60.00 unbound

Indexes:

50-year (1904–1953)	$12.00 cloth
10-year (1954–1963)	$3.50 paper
25-year (1964–1988)	$10.00 paper

Please inquire for shipping charges and discounts on back issues.

709 Philosophy Hall, Columbia University
New York, New York 10027

NEW FROM MIT

Studies in Contemporary German Social Thought

THE CRITIQUE OF POWER

Reflective Stages in a Critical Social Theory

Axel Honneth

Honneth's *Critique of Power* is a rich interpretation of the history of critical theory, which clarifies its central problems and emphasizes the "social" factors that should provide that theory with a normative and practical orientation.

376 pp $35 00

CIVIL SOCIETY AND POLITICAL THEORY

Jean L. Cohen and Andrew Arato

The authors contend that the concept of civil society articulates a contested terrain in the West that could become the primary locus for the expansion of democracy and rights

600 pp $45 00 (January 1992)

Now in paperback

THE NEW CONSERVATISM

Cultural Criticism and the Historians' Debate

Jürgen Habermas

308 pp $14 95 paper

THE STRUCTURAL TRANSFORMATION OF THE PUBLIC SPHERE

An Inquiry into a Category of Bourgeois Society

Jürgen Habermas

328 pp $14 95 paper

Zone Books

Now in paperback

THE NORMAL AND THE PATHOLOGICAL

Georges Canguilhem

327 pp $16 95 paper

"The Accursed Share is *a brilliant product of* [*Bataille's*] *loony-toons coupling of critical genres* "—Erik Davis, Voice Literary Supplement

THE ACCURSED SHARE

Volume 1

Georges Bataille

198 pp $12 95 paper

PERSPECTIVE AS SYMBOLIC FORM

Erwin Panofsky

Drawing on a massive body of learning Panofsky produces a type of "archaeology" of Western representation that far surpasses the usual scope of art historical studies

128 pp , 40 illus $22 95

[illegible]

THE MIT PRESS

55 Hayward Street, Cambridge, MA 02142

OF MIND & SELF

THE EMBODIED MIND

Cognitive Science and Human Experience

Francisco Varela, Evan Thompson, and Eleanor Rosch

"The Embodied Mind is *a thoroughly original integration of cognitive science, continental philosophy, and Buddhist thought, and in its transpersonal dimension, rather beautiful*"—Gordon G Globus, M D., *Professor of Psychiatry and Philosophy*, University of California, Irvine

"*Our concern is to open a space of possibilities in which the circulation between cognitive science and human experience can be fully appreciated and to foster the transformative possibilities of human experience in a scientific culture*"—*the authors*

The Embodied Mind corrects this imbalance within cognitive science by providing a deep and sophisticated treatment of the spontaneous and reflective dimensions of human experience.

300 pp , 34 illus $25 00

To order call toll-free 1-800-356-0343 Fax (617) 625-6660 MasterCard & VISA accepted

THE MIT PRESS

55 Hayward Street, Cambridge, MA 02142

THE JOURNAL OF PHILOSOPHY

is pleased to announce the publication of its

TWENTY-FIVE YEAR INDEX, 1964–1988

VOLUME LXI TO VOLUME LXXXV

AUTHORS, SUBJECTS, & REVIEWS

Supplement to its

FIFTY-YEAR INDEX, 1904–1953

TEN-YEAR SUPPLEMENT, 1954–1963

It is available for the rest of 1991 at special prices

25-year (1964–1988) $25 00 cloth/$10 00 paper

also available

50-year (1904–1953) $12 00 cloth 10-year (1954–1963) $3 50 paper

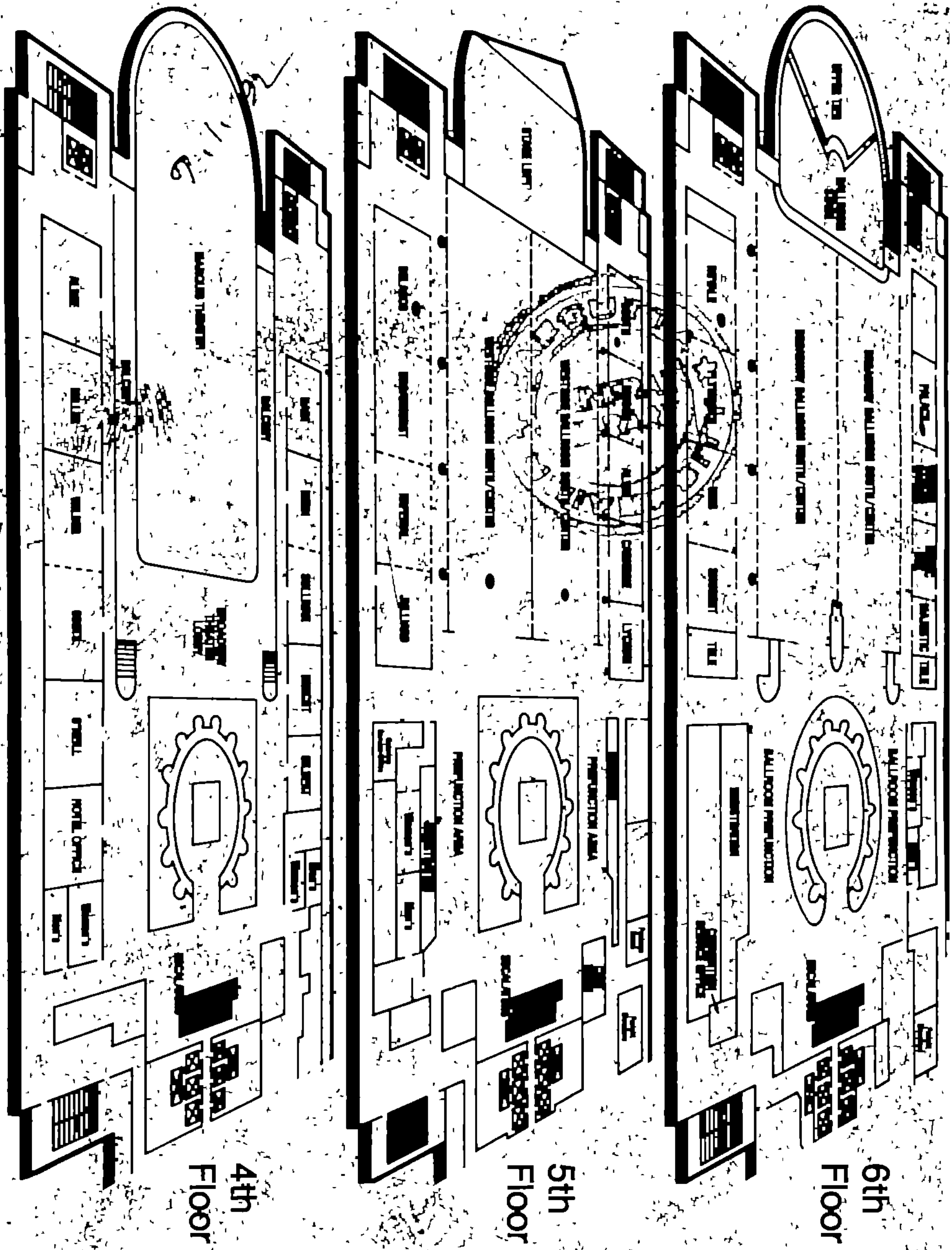
Shipping $5 00 (first)/$2 50 (each addl)

Please send a check or money order to

Journal of Philosophy
704 Philosophy Hall
Columbia University
New York, New York 10027

Prices are in effect through December 31, 1991 Orders and payment must be received by then

AMERICAN PHILOSOPHICAL ASSOCIATION, EASTERN DIVISION
88TH ANNUAL MEETING, DECEMBER 27–30, 1991
MARRIOTT MARQUIS HOTEL, NEW YORK



7th FLOOR: DUFFY/COLUMBIA and GRAMERCY/HERALD/SOHO ROOMS
9th FLOOR: SALONS A, B, & C; CANTOR/JOLSON and BARRYMORE ROOMS

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NUMBER 11
NOVEMBER 1991

Special Issue, Eighty-eighth Annual Meeting
American Philosophical Association, Eastern Division

page

Symposium: The Epistemology of Modality

607 *Plenitude of Possible Structures* Phillip Bricker
620 *The Plenitude of Structures and Scarcity of Possibilities* Joseph Almog

Symposium: Aquinas on the Veracity of the Intellect

623 *Aquinas on the Veracity of the Intellect* John Jenkins

Symposium: Feminist Critiques of Equality

633 *Equalities* Martha Minow
645 *In Whose Different Voice?* Eva Feder Kittay

Symposium: Can Biological Teleology be Naturalized?

647 *Can Biological Teleology Be Naturalized?* Mark Bedau
656 *Naturalism and Teleology* Mohan Matthen

Symposium: Knowledge, Power, and Science

658 *The Dynamics of Power and Knowledge in Science* Joseph Rouse
666 *Multiplying Subjects and the Diffusion of Power* Helen E. Longino
675 *Socializing Knowledge* Philip Kitcher

Symposium: What Makes Life Worth Living?

677 *Why Bother: Is Life Worth Living?* John J. McDermott
684 *Questioning the Question of the Worth of Life* Karsten Harries
691 NOTICE ABOUT 1992 APA EASTERN DIVISION MEETING
691 NOTES AND NEWS

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1991 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices
Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NO 11, NOVEMBER 1991

PLENITUDE OF POSSIBLE STRUCTURES*

OUR chief concern is with actuality, with the way the world is. But inquiry into the actual may lead even to the farthest reaches of the possible. For example, to know what consequences follow from a supposition, we need to know what possibilities the supposition comprehends. Suppose that space is unbounded. Does it follow that space is infinite, as was once generally believed? The possibility of "curved" space demonstrates the opposite. Inquiry is driven by logic, and logical relations hold or fail to hold according to what is logically possible.

Whence come our beliefs about logical possibility? Typically, they derive from the analysis of particular nonlogical concepts. But sometimes we reason in accordance with general principles that are constitutive of logical possibility itself, principles to the effect that, if such and such is possible, then such and so must be possible as well. I shall call these *principles of plenitude;* I divide them into three sorts. First, there are principles that require a plenitude of recombinations. We reason according to such principles when we argue that it is logically possible for there to be a human head attached to the body of a horse. Second, there are principles that require a plenitude of possible contents. We reason according to such principles when we argue that some or all of the actual individuals and properties could be replaced by individuals and properties not of this world: alien individuals and properties. Finally, there are principles that require a plenitude of possible structures. We reason according to such principles when we argue that, if it is logically possible for there to be four or five spatial dimensions, then it is logically possi-

---

* To be presented in an APA symposium on the Epistemology of Modality, December 30. Joseph Almog will comment. See this JOURNAL, this issue, 620–2, for his contribution

0022-362X/91/8811/607–19

ble for there to be seventeen, or seventeen thousand. These three sorts of plenitude, taken together, delimit the scope of the possible.

In this paper, I consider only plenitude of possible structures. I take it that there are structures we know to be logically possible, for example, the three-dimensional Euclidean and non-Euclidean spaces of constant curvature. My goal is to uncover the source of that knowledge, and thereby to combat skepticism about modality without appealing to any mysterious faculty of modal intuition. On my account, our knowledge starts from our theorizing about the actual world and is extended, in accordance with the demands of plenitude, by the results of mathematics. I develop and defend a principle of plenitude for structures, and motivate the principle pragmatically by way of the role that logical possibility plays in our inquiry into the world.

First, some preliminary points. A structure is logically possible, on my usage, only if there are or could be concrete entities that instantiate that structure, that is, only if the structure is instantiated by (some or all) of the concrete inhabitants of some possible world.[1] A structure is instantiated by a plurality of inhabitants of a world in virtue of their *natural* properties and relations; otherwise, structures would be too easily instantiated, since instantiation would depend only upon cardinality.[2] I assume that structures exist as abstract entities of some sort. If something more definite is wanted, structures may be represented set-theoretically in ways familiar from model theory for first-order languages.

I shall focus in what follows upon spatial and spatiotemporal structures. Not because I think these are the only sorts of structure to which plenitude applies: worlds have a pattern of instantiation of nonspatiotemporal natural properties and relations; and perhaps some worlds have irreducible causal, or nomological, or probabilistic structure. I focus upon spatial and spatiotemporal structures because they provide substantive examples upon which there is some initial agreement as to possibility.[3]

---

[1] Some authors use 'metaphysical possibility' for what I call 'logical possibility', and reserve 'logical possibility' for some (prima facie) weaker notion of "mathematical" (or "conceptual") possibility: a putative abstract entity is mathematically possible, roughly, if it can consistently be posited to exist. All structures are possible in this sense.

[2] Not every class of entities at a world is the extension of a *natural* property; belonging to the extension of a natural property may be a matter of shared universals, or duplicate tropes, or primitive naturalness applied to classes of *possibilia;* I need not decide that here. For discussion and comparison of these views, see David Lewis, *On the Plurality of Worlds* (New York: Blackwell, 1986), pp. 59–69.

[3] Beware. I normally use 'space' and 'spatial structure' interchangeably to refer to a "mathematical" entity; but 'space' also has a physical interpretation. Thus,

I

There is one more piece of business before turning to plenitude of structures. A principle of plenitude for structures, on my account, does not by itself determine which structures are possible. It serves rather as a principle of inference for modal reasoning: given that these initial structures are possible, these other structures are possible as well. The possibility of the initial structures must be believed on independent grounds. What might these be?

Consider Newtonian space-time: any two events have an absolute spatial and an absolute temporal separation. I assume we all believe that Newtonian space-time is logically possible. But, thanks to Einstein, we no longer believe it is actual, or even compatible with the actual laws. This suggests that logical possibility is required to encompass, not only actuality and nomological possibility, but our theorizing about actuality and nomological possibility as well. I propose:

> (B) We have warranted belief that a structure is logically possible if that structure plays, or has played, an explanatory role in our theorizing about the actual world.

A number of comments are in order. (1) Condition (B) makes warranted belief about logical possibility relative to history and to a community of theorizers, as it should; it does not make logical possibility itself relative. (2) As the case of Newtonian space-time suggests, the historical relativity is asymmetric: the structures believed with warrant to be possible by a community only increase over time. (3) If a bad theory takes hold in a community, positing gratuitous structure that explains nothing, condition (B) does not apply. To 'play an explanatory role' is not just to be *taken* to play an explanatory role by the community. The structure must have genuine explanatory power.[4] (4) Although it is primarily *scientific* theorizing about the world that I have in mind, (B) is not thus restricted. Philosophers and theologians have posited explanatory structures in

---

when I say that Euclidean space is instantiated at a world, I do not thereby say that physical space at the world is Euclidean. The latter requires that the structure, Euclidean space, be instantiated by the right entities (e g., all the points of physical space), and perhaps also in virtue of the right natural relations (e.g., the distance relation at the world).

[4] Perhaps even Newtonian space-time fails this test due to its gratuitous positing of absolute rest; in which case only so-called Neo-Newtonian, or Galilean, space-time, could be warranted by (B). The possibility of Newtonian space-time would then be derived from plenitude. For the distinction between Newtonian and Galilean space-time, and a discussion of the explanatory adequacy of spatiotemporal structures, see Michael Friedman, *Foundations of Space-Time Theories* (Princeton: University Press, 1983), pp. 71–92, 236–63.

their theorizing about the world. However much we mistrust their speculations, we should not exclude these structures without cause. We can eliminate bad philosophy or theology in the same way we eliminate bad science: by requiring genuine explanatory power. (5) It is enough for a theory to be seriously considered by a community; it need not ever be believed. Belief in the possibility of Lobachevskian space (of very small negative curvature) is warranted by (B), because it was seriously considered (in the nineteenth century) whether measurements of stellar parallax supported the Euclidean or Lobachevskian theory of space. (6) Whenever a structure is instantiated at a world, so are all its substructures. For example, a world at which three-dimensional Euclidean space is instantiated is also a world at which one- and two-dimensional Euclidean space is instantiated. Thus, warranted belief in the possibility of a structure passes to all of its substructures. For convenience, I shall interpret 'plays an explanatory role in our theorizing' in such a way that, whenever a structure plays such a role, all of its substructures do so as well.

II

I turn now to plenitude. There are structures we believe possible that neither play, nor have played, any explanatory role in our theorizing. We believe them possible, I suppose, because we believe that the space of logical possibilities must be "filled out" or "completed" in some nonarbitrary way. But what counts as arbitrary here? Can these constraints on logical space be made more precise?

As a first try, we might take the intuitive idea underlying plenitude to be that "there are no gaps in logical space."[5] But what constitutes a gap? Suppose that there are worlds with Euclidean space of six dimensions, and worlds with Euclidean space of eight dimensions, but none with Euclidean space of seven dimensions. Would that be a violation of plenitude, a gap in logical space?

It would, indeed; but one must be cautious in giving the reason. Six-sided regular polyhedra (cubes) are logically possible, as are eight-sided regular polyhedra (octohedra), but not seven-sided regular polyhedra. Yet that does not constitute a gap in logical space. Wherein lies the difference? There *is* a gap in the first case, because mathematical generalizations of three-dimensional Euclidean space to higher dimensions include a seven-dimensional space whenever they include six- and eight-dimensional spaces; and they provide a natural ordering of the spaces according to which the seven-dimensional space falls between the other two. There is *no* gap in the second case, because mathematics teaches us that a seven-sided regular polyhedron is a contradiction in terms; so in going from six-sided to eight-sided, nothing has been left out. In sum, mathematics provides the backdrop of structures and the natural orderings on structures, without which the notion of a gap in logical space would make no sense.

---

[5] From Lewis, p. 86

It is not enough, however, to rule out gaps in logical space; plenitude demands that logical space contain no arbitrary or unnatural boundaries. Suppose that Euclidean spaces of all dimensions up to six were logically possible, but none of greater dimension. That, too, would be a violation of plenitude. The mathematical generalization of three-dimensional Euclidean space to four-, five-, and six-dimensional Euclidean space applies, *mutatis mutandis,* to all finite dimensions; there is no natural stopping point among the finite-dimensional spaces. To allow that some but not all finite-dimensional Euclidean spaces are logically possible would be to assign an unnatural boundary to logical space.

The idea that logical space contain no unnatural boundaries can be taken to supercede and clarify the idea that it contain no gaps. A gap in logical space is formed by two boundaries, one from either side. Call a gap *natural* if both its boundaries are natural; *unnatural* otherwise. A prohibition on unnatural boundaries entails a prohibition on unnatural gaps; natural gaps in logical space, if any there be, need not be a violation of plenitude.

It should be apparent by now that an account of plenitude must rely heavily on a notion of naturalness (or some equivalent). I shall assume that naturalness applies to classes generally, and, in particular, to classes of structures. Talk of natural boundaries in logical space is easily translated into talk of natural classes: any class of logically possible structures determines a boundary in logical space; the boundary is *natural* just in case the class is natural, or is a union of natural classes. Although I have no analysis of naturalness to offer, some words of clarification and illustration are in order.

Naturalness applies both to classes of physical entities and to classes of mathematical entities. In either case, what the natural classes are is not determined by us: it is a matter of objective, noncontingent fact. Examples of natural classes of mathematical entities include: the natural numbers, the real numbers, the ordinal numbers, recursive functions of natural numbers, continuous functions of real numbers. Examples of natural classes of mathematical structures include: groups, vector spaces, topological spaces, Euclidean spaces. Each of these natural classes serves as the principle object of study for some major area of mathematics. If a working criterion for naturalness is wanted, we have here, at least, a sufficient condition. That is not to say, however, that the above-mentioned classes are natural *because* mathematicians have chosen to study them. Rather, mathematicians have chosen to study them, I take it, in part because they are natural classes.

Although I shall speak of classes simply as natural or unnatural, it is clear that naturalness is a matter of degree. The odd natural numbers do not form a natural class in the sense here intended: the study of odd number theory, as opposed to number theory, would be a largely fruitless endeavor. But the odd numbers deviate from naturalness less than the numbers that are odd up to a hundred and even thereafter; and these numbers in turn deviate from naturalness less than some really gruesome class of numbers not even definable within elementary arithmetic. For what follows, I need to assume that classes of structures may be *perfectly* natural, that there is a greatest degree of naturalness; when I say 'natural', I mean 'perfectly natural'.

Naturalness itself imposes a structure on the classes of structures. Some assumptions about this structure will be needed below. I assume that the natural classes exhaust the class of all structures, that is, that every structure belongs to some natural class. I assume that the natural classes are not closed under unions or complements; though it is plausible that they are closed under intersections, that is controversial, and I shall not assume it in what follows. Finally, I assume that the class of all structures is *not* a natural class, on grounds of heterogeneity; but the formulations below could easily be revised to accommodate the contrary judgment.

III

With the notion of naturalness of classes in hand, I turn to formulations of a principle of plenitude for structures. The easiest way to meet the demand that there be no unnatural boundaries is to draw no boundaries at all:

(P1) Every structure is a logically possible structure.

I find (P1) attractive as a principle of plenitude for structures. For one thing, it provides an exceedingly simple account. Once (P1) is accepted, (B) becomes superfluous; mathematics alone—perhaps, mathematical logic alone—determines which structures are possible. Moreover, though the notion of naturalness may play a role in motivating (P1), it plays no role in its formulation. Unfortunately, (P1) goes far beyond anything demanded by the idea that logical space be characterizable in a nonarbitrary way. Perhaps (P1) could be defended by way of the benefits it confers upon our total theory.

In any case, I shall here remain agnostic toward (P1), and go on to develop a (somewhat) more conservative principle that is capable of a stronger defense.

There is another simple way to meet the demand that the space of possible structures contain no unnatural boundaries:

(P2) The class of logically possible structures is a natural class.

(P2) constrains the shape of logical space. It does not by itself tell us whether any particular structure is logically possible. But when combined with (B), it may support inferences to the possibility of particular structures. Thus, let *B* be the class of structures warranted by (B). Any structure that belongs to every natural class of structures which includes *B* is warranted by (P2). (I say a structure is *warranted,* for short, if belief in its logical possibility is warranted.) For example, suppose that *B* contained only the Euclidean spaces of one-, two-, and three-dimensions; then (P2) would warrant the other finite-dimensional Euclidean spaces.

(P2) will not do as a principle of plenitude for structures, however: it is both too strong and too weak. To see that it is too strong, consider the class of logically possible spatiotemporal structures. I take it we believe, based upon (B), that this class includes both continuous and discrete space-times, but I do not believe that any natural class encompasses them both; the mathematics of continuity and the mathematics of discreteness have little in common. Thus, *B* is not included in any natural class, making the acceptance of (P2) incompatible with (B).

A solution is not far to seek. Although the class of possible space-times is not a natural class, it is a *union* of natural classes; we call them all "space-times" not because they form a natural mathematical kind, but because of some looser family resemblance. This suggests that we weaken (P2) as follows:

(P3) The class of logically possible structures is a union of natural classes.

(P3) still constrains the shape of logical space, assuming, at any rate, that singletons are not in general natural classes. But (P3) is genuinely weaker than (P2) because the natural classes are not closed under unions. Moreover, when combined with (B), it still supports inferences to the possibility of particular structures: given a structure *b* in *B*, (P3) warrants any structure that belongs to every natural class containing *b*. Finally, (P3) is still sufficiently strong to guarantee that the space of possible structures contain no unnatural boundaries.

Nevertheless, I think (P3) is too weak in at least two ways. And if I am right, the condition that logical space contain no unnatural boundaries cannot be sufficient for plenitude. First, there is a problem of *crosswise generalizations*. Suppose that there are two natural ways of generalizing from a structure *b* in *B*, resulting in two natural classes containing *b*. If these generalizations cut crosswise, they may have only the structure *b* in common; in which case, no inference from the possibility of *b* to the possibility of any of the structures that generalize *b* will be supported by (P3). Consider this example. Suppose again that three-dimensional Euclidean space is one of the structures in *B*. One can generalize the number of dimensions to any finite value while keeping the space Euclidean, or generalize the curvature to any constant negative or positive value while keeping the space three-dimensional. Both generalizations, it seems to me, result in natural classes of spaces. It is compatible with (P3) that the spaces from only one of these classes be possible. But that is too weak. I think we have grounds to infer that *all* the spaces in question are possible, grounds that (P3) fails to capture. (P3) allows crosswise generalizations in effect to cancel each other out, without consequence.

One might simply concede that crosswise generalizations on a single structure *b* cancel one another *unless* there are other structures in *B* that, together with *b*, support inferences to the structures that generalize *b*. Thus, plenitude of structures demands that all finite-dimensional Euclidean spaces be possible only because *B* contains, in addition to the three-dimensional Euclidean space, the one-, and two-dimensional Euclidean spaces; and any natural class containing these three spaces contains all finite-dimensional spaces. (Similarly, all three-dimensional spaces of constant curvature are possible because *B* contains three-dimensional spaces of (very small) negative and positive constant curvature.) This suggests it might suffice to enhance (P3) as follows:

> (P4) The class of logically possible structures is a union of natural classes. Moreover, suppose *S* is a class of logically possible structures that is included in some natural class. Any structure that belongs to every natural class of structures that includes *S* is logically possible.

(P4) falls midway in strength between (P2) and (P3): unlike (P3), it permits inferences from *classes* of structures, not just from *single* structures; but unlike (P2), it does not require that every class of possible structures be included in some natural class.

Is (P4) strong enough to capture plenitude of structures? I think not. For (P4) as well as (P3), there is a problem of *nested generalizations*. Consider the supposition that there are possible Euclidean spaces with any finite number of dimensions, but no possible Euclidean spaces with infinitely many dimensions. This supposition posits no unnatural boundaries in logical space: the class of finite-dimensional Euclidean spaces is a natural class, an appropriate object of study in mathematics. Thus, the supposition violates neither (P2), (P3), nor (P4). But I claim it is a violation of plenitude nonetheless. The natural generalization of one-, two-, and three-dimensional Euclidean space to other finite dimensions can itself be naturally extended into the infinite. For example, there is a natural generalization of the Euclidean metric to spaces of continuum-many dimensions which makes use of the way that integration generalizes finite summation.[6] Assuming that the Euclidean spaces in B are all finite-dimensional, it follows that they are included in at least two natural classes, one a subclass of the other. (P4) provides no grounds for inferring that any space contained only in the larger of the two subclasses—that is, any infinite-dimensional Euclidean space—is logically possible. But on what grounds does plenitude differentiate between the possibility, say, of a seventeen-dimensional Euclidean space, and the possibility of an infinite-dimensional Euclidean space? What does the size of a spatial structure have to do with the possibility of its instantiation?

One might reply: the seventeen-dimensional space is *closer* to the spaces in B than any infinite-dimensional space, according to the natural ordering of structures. But this reply is incompatible, at least in spirit, with the all-or-nothing approach to logical possibility taken by (P2) through (P4). If a relation of closeness to the structures in B is what differentiates the finite- and infinite-dimensional spaces with respect to possibility, it becomes an utter mystery why a space of seventeen-thousand dimensions should be no less possible than a space of seventeen. The reply in question leads inevitably, I think, to the view that logical possibility is a matter of degree, in which case logical implication becomes a matter of degree as well. That is a truly radical view; I do not reject it out of hand, but it will not be considered further in this paper. I know of no other grounds for favoring the finite-dimensional over the infinite-dimensional Euclidean spaces. I conclude that any principle of plenitude that warrants belief in the possibility of the former, must warrant belief in the possibility of the latter. (P4) fails this test.

[6] A standard example. Let the points of the space be the continuous real-valued functions defined on the real interval [0, 1]. Define the distance between two points, f and g, to be: $\sqrt{\int_0^1 (g(x) - f(x))^2 dx}$

The same conclusion can be reached by a slightly different route. Suppose again that plenitude requires that there be no arbitrariness in logical space. One way for logical space to be arbitrary, I have said, is to have an unnatural boundary, that is, to not be a union of natural classes. But there is another way. Consider a nested sequence of natural classes representing more and more high-powered generalizations of some structures in *B*; suppose that any member of *B* occurs in the first member of the sequence or in no member at all; suppose further that any natural class that includes every class in the sequence is itself a member of the sequence. If *Z* is the union of all classes in the sequence, then *Z* contains all the structures that are candidates for logical possibility in virtue of the mathematical generalizations of the structures in question in *B*. Now, (P4) permits any division of *Z* into possible and not possible, so long as the possible structures form a natural class (and include the given structures in *B*). But it would be arbitrary for the boundary of logical space to follow one such division over any other. The only way to avoid such arbitrariness in logical space is to impose no division of *Z*. This suggests the following principle of plenitude:

> (P5) Suppose *s* is a logically possible structure. Any structure that belongs to any natural class of structures containing *s* is logically possible.

(P5) substantially strengthens (P4). When combined with (B), it supports inferences to the possibility of spaces of any infinite dimensionality, as long as those spaces arise from a natural mathematical generalization of ordinary Euclidean space.

I wish I could in good conscience stop here; but a complication remains. There is a problem of *overhasty generalization*. Consider one-dimensional Euclidean space; that is, the structure of the real numbers with the usual metric: distance $(x, y) = |x - y|$. Is there any natural process of generalization that, when given *only* this structure as input, gives the finite-dimensional Euclidean spaces as output? I think not. The fundamental form of the Euclidean metric —being the square root of a sum of squares—plays no role in the one-dimensional case. Granted, one-dimensional Euclidean space is a *special case* of finite-dimensional Euclidean space; but it is too trivial a special case to support a generalization to higher dimensions. This leads to a problem with (P5). Given the possibility of only the one-dimensional Euclidean space, (P5) supports the inference to the possibility of all the finite-dimensional Euclidean spaces. That inference seems just as overhasty as the generalization upon which it is based.

There is an easy fix that should be resisted. We could say that plenitude of structures only supports inferences based upon generalizations involving *two or more* structures. But that fails to get to the heart of the problem. Natural generalizations can, I think, be based upon a single structure if that structure is not a trivial or degenerate case of the generalization; perhaps three-dimensional Euclidean space is an example. On the other hand, two structures may be no better than one, if both structures are trivial cases of the generalization in question. The number of structures needed to support a generalization is relative both to the type of generalization and to the particular structures chosen; it cannot be specified, once and for all, in advance.

I see no choice, then, but to conclude that the notion of natural class is not by itself sufficient for formulating a principle of plenitude for structures; we need a *relation* that holds between a class of structures and those classes of structures that are *natural generalizations* of it. A natural generalization of a class of structures is always a natural class; but a natural class need not be a natural generalization of all of its subclasses. Switching from natural classes to natural generalizations transforms (P5) into:

> (PS) *Plenitude of Structures.* Suppose S is a class of logically possible structures. Any structure belonging to any natural generalization of S is logically possible.

This is the principle of plenitude for structures which I accept. It shares all the virtues of (P5): the logical space of possible structures has no unnatural boundaries, nor arbitrariness in the way boundaries are set.

IV

Thus far I have assumed without argument that logical space should have natural boundaries set in a nonarbitrary way. Can this assumption itself be defended? I think it can. I take it to be constitutive of logical possibility that it provide a suitable framework for our inquiry into the actual world; whoever denied this could not mean what I do by 'logical possibility'. Our inquiry into the actual world involves concepts—such as space, time, and space-time—that have meaningful application beyond the actual world, indeed, beyond the nomologically and the doxastically possible worlds. Since part of that inquiry is inquiry into the nature of these concepts and their logical interrelations, logical possibility must extend at least as far as the meaningful application of these concepts.

Consider the question with which I began this paper: If (physical) space is unbounded, must it also be infinite in extent? Suppose the question had been asked in the eighteenth century, prior to the discovery of non-Euclidean geometry. I think the answer would have been "no" even then: 'space' did not then mean 'Euclidean space', any more than it does now. Thus, questions about the world that might well have been asked in the eighteenth century could only have been answered in the light of mathematical generalizations that were then unknown. The situation is no different today. We do not know in advance which mathematical generalizations of our concepts will turn out to be relevant to our inquiry.[7] If the class of logically possible structures includes some but not all of these generalizations, as is allowed by (P2) through (P4), then logical possibility may be unfit to provide a logical framework for our inquiry into the world. In order to ensure that no relevant structure is left out of logical space, we need to posit a plenitude of possible structures, we need the space of possible structures to be filled out in a nonarbitrary way.

The role that logical possibility plays in inquiry can motivate and justify both (B) and (PS); does it also support (P1), that every structure is logically possible? No; logical possibility must be broad enough to accommodate inquiry into matters of *contingent* truth, not matters of *necessary* truth. I do not require, nor is it customary to require, that logical possibility provide a framework for mathematics. If a structure does not belong to any mathematical generalization of any actual structure, or of any structure warranted by (B), then it is logically irrelevant to inquiry into the actual world.[8] It could safely be excluded from logical space.

V

There is no space to further illustrate my account, or to compare it with what others have said.[9] I conclude by summarizing the implications of my account for the epistemology of modality. I have attempted to steer a course between the Scylla of modal skepticism and the Charybdis of an obscurantist modal epistemology. The skeptic I have in mind holds that our only grounds for belief in the possibility of structures are the theoretical and explanatory grounds embodied in (B). Such skepticism is belied by ordinary practice, by our ordinary ways of thinking about modality. I take it our role as philosophers is not to challenge ordinary practice—except perhaps in rare cases—but to attempt to account for ordinary practice in a systematic way. I have developed and defended a principle of plenitude, (PS), that I think adequately explains and locates the source of our belief in a plenitude of possible structures. It warrants belief in the possibility of structures that are not ordinarily thought to be possible; but these structures are not ordinarily thought to be possible, I think, only because they are not ordinarily thought of at all. My account does not purport to eliminate all ignorance as to which structures are logically possible. If a structure is not warranted by (PS) together with (B), it may or may not be logically possible, for all I have said; an absence of warranted belief does not warrant belief in the contrary.[10] Moreover, there is ignorance associated with the application of (B) and (PS): ignorance as to which structures are explanatorily adequate to actual phenomena translates into ignorance as to which structures are possible; as does ignorance as to the mathematical generalizations of structures. But although it may sometimes be unclear how my account applies in a particular case, the general grounds of our beliefs are made clear. When we infer that some structure is possible using (B) and (PS), we are guided by science (broadly construed) and by mathematics, not by some mysterious faculty of modal intuition. Nor is any such faculty needed to motivate or defend (B) and (PS). They are motivated and defended, not by modal intuition, but by what we require a theory of modality to do. And there need be nothing obscure about that.

---

[7] Actually, I hold something stronger, that we know in advance that every generalization is logically relevant, so long as it is compatible with whatever necessary conditions we place on the concept. But that depends upon a theory of content for concepts that I shall not defend here.

[8] Of course, it may be psychologically relevant by suggesting analogies, serving as a heuristic tool, and so on.

[9] Robert Adams appears to reject what I call plenitude of structures altogether. He holds that only the sort of considerations embodied in (B) can warrant belief in the possibility of structures. See "Presumption and the Necessary Existence of God," *Nous*, XXII (1988), 19–32. Lewis believes in a plenitude of possible (spatiotemporal) structures, but his account is based upon some principle weaker than (PS). He would reject (PS) because it leads to there being no *set* of all possible worlds. (I accept the consequence.) Cf. Lewis, pp. 104–8.

PHILLIP BRICKER

University of Massachusetts/Amherst

---

[10] My account is compatible with the view that *only* the structures warranted by (B) and (PS) are possible. But I would reject that view on grounds of parochialism; it would allow features of *our* inquiry, contingent and accidental though they be, to delimit the scope of the possible.

THE PLENITUDE OF STRUCTURES AND SCARCITY OF POSSIBILITIES*

Given some *ur*-possibilities, what further possibilities are projectible? Phillip Bricker's interesting answer involves what I shall call a ***projection schema:***

> (PS) If *S* is a class of logically possible structures, any natural generalization of *S* is logically possible.[1]

Such projection principles shed light on what there is as well as how we are to know that it is. Bricker is very much a cartographer of the possible; but the end of the tunnel for him is modal ***knowledge:*** "to steer a course between—the scylla of modal skepticism and charybdis of an obscurantist modal epistemology." Contemporary epistemology seems spellbound by the Greek world, what with tragic "dilemmas" and Antigonian choices galore. The present writer would rather keep to his Dionysian lifestyle of a naive caveman; his elective affinities consist of focusing on (PS) as an ***existence*** principle.[2]

Bricker reads (PS) as a "principle of plenitude." Of what? Primarily of "structures," derivatively of "possibilities." Plenitude may be true (intrinsic) of structures, not of possibilities. Four kinds of (counter) examples concerning (PS) are to be discussed.

First, consider "general (Henkin) models," so dubbed by their originator because they were natural generalizations of the rather specific standard originals. Can we project from them possibilities? We had better not. The (arithmetic, set theoretic, etc.) truths of the standard structure are necessary truths, articulating *what* the concerned items are. The consistent falsehood of some of them, witnessed by an appropriate general model, should be taken for what it

---

* Abstract of a paper to be read in an APA symposium on the Epistemology of Modality, December 29, commenting on a paper by Phillip Bricker. See this JOURNAL, this issue, 607–19, for his contribution.

[1] The reader is to keep in mind that Bricker does not mean by 'logically possible' mere "(semantic) consistency" (with some underlying first-order theory). For the meaning(s) he does have in mind, see his paper above.

[2] Also to be varied is Bricker's diet: many of his examples come from (differential) geometry. The present writer's knowledge of the (meta) mathematics involved is insufficient.

0022-362X/91/8811/620–2

is—the witnessing of consistency, not for another thing—a genuine possibility.[3]

A second example comes from a "more standard," perhaps the most natural producer of new structures out of old—ultraproducts (powers).[4] Again, natural metamathematics is no guarantor of sound metaphysics—ultrapowers are ultrapowers, not ultrapossibilities. A simple example: taking ω copies of the standard integers (under a suitable ultrafilter), one can get eventually a structure with "infinite" integers, e.g., a number n^* not reached by finitely many iterations from 0. A possibility it is not. Neither on the ambitious, "wide scope," reading: (1) there exists a natural number of which it is possible that it takes infinitely many steps to reach; nor even on the weaker reading: (2) it is possible that there is a natural number that takes infinitely many steps to reach.

A third example arises with theories not taken to allow the standard/nonstandard model distinction. It has been argued that the class of first-order predicate calculus ("free") models with the empty domain included is a natural generalization of the classical multiplicity. So it has also been argued on behalf of Boolean-valued models, assignments of (truth) values in an arbitrary Boolean algebra rather than exclusively of 0 and 1. Let these claims stand. Should this be

---

[3] I shall consider three examples of nonmodally projectible generalized structures: (i) Saul Kripke's (with a twist due to Steven Thomason) generalized partial truth tables for propositional S5. (PS) would have us cut the branch we are sitting on—it is the generalized tables that bar (the logical validity of) otherwise well-grounded . . . *ur*-possibilities; (ii) Generalized structures for a second-order modal language (with propositional quantifiers) that deprive us of the (necessary) existence of a truth recording everything that is the case; (iii) Generalized structures with infinitely descending sequences of the ϵ-relation. Time permitting, I shall want to consider a stronger brand of examples—theories (one in propositional tense logic, the other W. O. Quine's handling of arithmetic in "New Foundations") with no standard models at all whose (possible) consistency *has* to rely on general models. With such models the failure of projectibility of generalized structures is quite vivid.

[4] To keep technical foam to a minimum, consult for the definition of the notion(s) and the extension alluded to below of a Frechet filter to a maximal filter, Bell and Machover, *A Course in Mathematical Logic* (Dordrecht: North Holland, 1977), pp. 136–9 (especially p 138) and pp. 176–82. This much concerns models for arithmetic. A crisp exposition of nonstandard models of real-number theory, relying on related techniques, is offered later in the aforementioned book. In view of philosophical objections to the possibility of "infinitely small quantities," many might think that there lies here a striking case of nonmodally projectible class of generalized structures. For the present writer, when it comes to judgments of modal (im)possibility, it is either that obscurantist totem and taboo, modal intuition, or silence. In this case, the latter.

CENTRAL LIBRARY

sufficient to convince us, in the mantle of metaphysicians, that the world might have been vacuous (what does this mean)? That such a "Boolean assignment" represents, let alone is, a complete way the world itself might have been?

Fourth, to compare the very idea of "structure" and "possibility," consider the following generation principles of sets (structures) that many would vouch intrinsic to the notions concerned: (S1) the (set theoretic) idea that, if a (certain kind of natural) model of set theory satisfies a condition, a set already does; (S2) the (model, nonset, theoretic) idea that strong closure operations breed, when applied to given structures (e.g., the above-cited product), yet further structures and plenty at that. Either with sets or structures, it is ever more. Are there such counterpart principles with "possibility" substituting for "set" ("structure")?

One hopes not for (S2), where aggregation of all possibilities (gotten up to a stage) into a "super possibility" could misfire. Just pairing a couple of them might be embarrassing enough. As for (S1), as mentioned, rather than maximize possibilities, by way of consistency-verifying natural generalizations, we are bent on minimizing: the truths of the ground standard structure which pertain to the item's whatness delimit the modal horizon. With possibilities, less is more.

JOSEPH ALMOG

University of California/Los Angeles

AQUINAS ON THE VERACITY OF THE INTELLECT*

"The intellect," Aquinas frequently writes, "cannot be false."[1] He is not saying, of course, that we are not capable of false judgments—it is as obvious to him as it is to us that we are; he is saying, rather, that the intellect in its first operation, in which it forms the ideas or concepts that are to be used in making judgments, cannot be false. Even so, the claim seems bold and surprising, for inaccuracies in conceptualizations seem common enough. What is Aquinas's point?

Let us label this claim of Aquinas, whatever it amounts to, the *doctrine of the indefectibility of the intellect*, or just the *indefectibility doctrine*. At least two readings of the indefectibility doctrine can be distinguished: a weaker reading, which claims that the indefectibility doctrine is compatible with inadequacies in initial conceptualizations of natural kinds, such that, in light of further inquiry and judgments, the earlier conceptual scheme will be shown to be in need of revision; and a stronger reading, which denies such a compatibility. The weak reading has won general acceptance by prominent twentieth-century scholars of Aquinas: Etienne Gilson,[2] Bernard Lonergan,[3] and Alasdair MacIntyre[4] have all adopted some version of it. In this paper, however, I shall argue for a stronger alternative.

I THE INDEFECTIBILITY DOCTRINE

According to Aquinas, human cognition in its natural state begins with and remains tied to sense perception: there must be a *conversio*

---

* To be presented in an APA symposium on Aquinas on the Veracity of the Intellect, December 29. Eleanor Stump will comment. Anthony Kenny, Brian Davies, Ralph McInerny, David Burrell, Fred Suppe, and Michael Buckley were all most helpful in commenting on various versions of this paper.

[1] This claim is most extensively discussed in Aquinas's *In Aristotelis librum de anima commentarium*, III.xi (hereafter this work will be referred to as *In de anima*) and in *Summa theologiae*, I.85.6 (hereafter ST). Among the other places one finds it are: ST, I.17.3 and 58.5; *Summa contra gentiles*, I.59 and III.108 (hereafter SCG); *Quaestiones disputatae de veritate*, 1.12 (hereafter *De Ver.*); *In Aristotelis libros peri hermineias*, *lectio* iii (hereafter *In periherm.*); and *In duodecim libros metaphysicorum Aristotelis expositio*, VI.iv.

All references are to the respective Marrietti editions, and all translations are mine

[2] *Thomistic Realism and the Critique of Knowledge*, M.A. Wauck, trans (San Francisco Ignatius, 1986), esp. pp. 199–203.

[3] *Verbum: Word and Idea in Aquinas* (Notre Dame: University Press, 1967), esp. ch. II.

[4] *First Principles, Final Ends and Contemporary Philosophical Issues* (Milwaukee: Marquette, 1990), pp 46–7

0022-362X/91/8811/623–32

ad phantasmata, a turning to sense images, not only for one to form an idea,[5] such as the idea of triangle, but also to engage in geometrical speculation about the properties of triangles. Yet human, abstract thought, which is of course beyond mere sense perception, also requires the intellect with its natural light.

Thomas distinguishes two operations of the intellect (*In periherm., prooemium.* I). In the first operation of the intellect (FOI), one grasps the essences, or quiddities, of things. There seem to be two stages in the FOI: in the first stage, the receptive intellect receives the intelligible forms that are illumined by the natural light of the intellect in *phantasmata;* having received this determination, in the second stage the agent intellect can form an abstract idea, a *ratio.*

In the second operation of the intellect (SOI), one engages in *compositio* and *divisio,* positive and negative judgments. In his commentary on Aristotle's *De Interpretatione,* Thomas distinguishes between two sorts of *compositio* (*In periherm.,* I.iii.26).[6] The first, which I shall call *predicative compositio,* is present in both the affirmative judgment that Socrates is sitting, and the negative judgment that Socrates is not sitting; it is also present in the wish and the command, as well as in entertaining the thought, that Socrates is sitting. It is simply that joining of one idea to another which is involved in each of these mental acts. In the second sort of *compositio,* on the other hand, what are "joined," as it were, are the mental synthesis and reality in an affirmation; that is, one not only forms a mental synthesis, but judges that there is some real composition corresponding to it. I shall call that second sort *judgmental compositio.* In its complement, *divisio,* one denies the existence of the corresponding composition in a negative judgment. *Compositio* and *divisio* are each forms of assent. Truth and falsity enter into judgmental *compositio* and *divisio* because individual things and their forms may or may not be as they are judged to be.

When Aquinas, using '*intellectus*' to refer to the intellect in its first operation, repeatedly tells us, "intellectus non potest esse falsus"—"the intellect cannot be false," there is one clear and uncontroversial sense in which Thomas intends this statement: in the FOI, an idea is formed, but there is no assertion about whether the form does or does not exist *in rebus,* in substances in the world. Truth and falsity only enter in the SOI, when there is a judgmental *compositio* or *divisio*—an affirmative or negative judgment. But, in the FOI, no false judgments are made because no judgments are made; and thus the FOI cannot be false (or true, for that matter) (*In de anima,* III.xi.760). So this is not an indefectibility which is due to being always correct or true, but, rather, to stopping short of judgmental *compositio* or *divisio,* in which thought and reality are "compared" (a "*comparatio,*" as Aquinas says) and assertion is made.

---

[5] I use 'idea' to translate '*ratio*', which is the nature or essence of the thing abstractly considered, with *materia communia* but without *materia signata.* Such *rationes* are the objects of abstract thought.

[6] These two sorts of Thomistic *compositio* have been distinguished and discussed by Anthony Kenny in *Action, Emotion and Will* (New York: Routledge & Kegan Paul, 1963), pp. 225–7

To say that the intellect is indefectible in this sense is, of course, not to say much; it is like saying that the Pope is infallible as long as he makes no claims about God or the world, and even the most anti-Papal Protestant could agree with that! But is there an epistemologically more interesting sense in which the FOI is not liable to error? In his commentary on book III of Aristotle's *De Anima,* Aquinas suggests there is.[7] There he writes that there are *two* senses in which the FOI is not liable to being false: "This intellection [in the first operation of the intellect] is with regard to that about which [the intellect] is not false: not only because what is non-complex is neither true nor false, but also because it is not deceived with regard to what it is (*quod quid est*), as is said below" (*In de anima,* III.xi.746). Later Aquinas further explains the sense in which the intellect cannot be false, and is indeed true, in this second way: he writes, "the intellect in understanding what is non-complex is true insofar as it is assimilated [*adaequatur*] to the thing understood" (*In de anima,* III.xi.761).

It seems, then, that Aquinas wants to say that the FOI is indefectible not only in the sense of being neither true nor false by stopping short of judgments about the world, but also, more positively, it is in some sense "always true."[8] For the sake of convenient reference, let us label the former doctrine the doctrine of the *infallibility of the intellect,* and the latter the doctrine of the *veracity of the intellect.*

---

[7] Some caution against attributing to Aquinas views expressed in his commentaries on Aristotle, for these "are so many expositions of the doctrine of Aristotle, not of what might be called his own philosophy"; Gilson, *History of Christian Philosophy in the Middle Ages* (London: Sheed & Ward, 1955), p. 376. My own position is closer to M.-D. Chenu's, who held that Aquinas intended to present Aristotle's views in the commentaries, but always with the intention to advance his own inquiry; so we can expect him to say therein if he thinks some philosophical claim is philosophically unjustified; *Introduction a L'Etude de Saint Thomas D'Aquin* (Montreal: Institut D'Etudes Medievales, 1950), p. 177.

Whatever one's general understanding of the commentaries, however, one can take Aristotle's views of the indefectibility of the intellect in Book III of the *De Anima*—as Aquinas understands these, of course—to be Aquinas's own; for Thomas, when writing not exegetically but on his own, regularly refers us to this part of the Aristotelian corpus. (See ST, I.58 5 & 85.6; SCG, I.59; *De Ver.,* 1 12.10.22.)

The sense of Aquinas's infallibility doctrine seems unproblematic; but what does Aquinas mean in claiming that the intellect is always veracious?

II

Insofar as they speak of veracity, advocates of the weaker reading seem to suggest that Aquinas's subject correctly grasps the content of her own ideas, and thus has a veracious grasp of certain analytic claims;[9] but, of course, this sort of veracity is compatible with the possibility that her ideas do not correspond to any real essences. Although I believe, for a variety of exegetical reasons, that this interpretation cannot stand, the limitation on the length of this paper prevents me from presenting all of them here. Thus in this section, I shall offer only some central considerations motivating a stronger reading.

In Aquinas's main argument for his claim that the FOI is always true, he compares this operation of the intellect to sense perception, arguing that the former is always veracious in a way analogous to that in which the latter is. According to the weaker reading of veracity sketched above, the FOI is veracious because it does not concern the relationship between thought and its objects in the world, but simply its own ideas. Thus, on this reading, the veracity of the intellect is due to the privileged access a subject has to her own ideas. As we move to sense perception, then, we would expect Thomas to focus upon claims whose truth do not depend upon the object of sense, but upon the subject's own internal sensings, to which she has a privileged access. Candidates for such claims would be those like the one expressed by the Chisholmian, "I am appeared redly to."[10] This claim is incorrigible because whether it is true depends only upon internal factors, not upon external objects of visual perception.

When we look at Thomas's arguments, however, we find he says something quite different. The argument is essentially the same in the commentary on the *De Anima* and in the *Summa,* but more fully presented in the latter work. There Thomas writes:

---

[8] "Intellectus . . . secundum quod intelligit quid est res, verus est semper . . ." (*In de anima,* III.xi 761).

[9] Lonergan writes that, in the FOI, "as yet one knows not the thing, but the idea of the thing" (*op. cit.,* 8). Insofar as the FOI is veracious, then, it would seem to be in regard to its knowledge of "the idea of the thing."

[10] Roderick Chisholm, *Theory of Knowledge,* Elizabeth and Monroe Beardsley, eds. (Englewood Cliffs: Prentice-Hall, 1966), p. 34.

> The senses are not deceived about their proper objects. For example, sight is not deceived concerning color—except when, as may happen, in an incidental way there is an impediment of the organ; as, for instance, in the case of feverish persons, the sense of taste judges sweet things to be bitter, because of the fact that the tongue is filled with bad humors (ST, I.85.6c).

Clearly, the veracity here cannot be because the reports in question are only of subjective seemings; these reports are just as true for someone with a fever, but Thomas says his perception of taste is not veracious. The veracity here is with respect to the proper object; in the case of sight, these are the colors of things in the visual field. Sight is veracious in the sense that by it, given that the organ is healthy and that there is sufficient light, one perceives the actual colors of things, sight's proper object.

In support of this claim, Aquinas offers a general argument that applies not only to sense perception, but to any cognitive potency:

> The reason for this [veracity of sense perception] is obvious. For, each potency, as such, receives an intrinsic, non-incidental ordination to its proper object (*per se ordinatur . . . ad proprium objectum*). But whatever has this sort of ordination is always the same. Thus, as long as the potency remains intact, judgements concerning its proper object are not defective (*non deficit*) (ST, I.85.6c).

The presupposition of this argument seems to be that cognitive potencies are directed to gathering information about the world. The critical feature of potencies that are always veracious is that they always act in the same manner, barring the loss of the potency, as when sight or hearing is lost through damage to the corporeal organ. Apart from such impediments, error enters only in the case of potencies that are under the control of the will, and can act sometimes in one way, sometimes in another. Sense perception acting always in the same way invariably achieves its cognitive ends.

Applying this to the intellect, Thomas writes:

> The proper object of the intellect is the quiddity of a thing. Thus, the intellect is not deceived concerning the quiddity of a thing (given that we are speaking of what is intrinsic and non-incidental to [the quiddity]) (ST, I.85.6c).

Thus, it is not, as the weaker reading suggests, that the FOI is veracious in correctly grasping its own ideas. Rather, it correctly apprehends the quiddities or essences of those things to which it has access through sense perception.

These passages suggest that Aquinas's doctrine of the veracity of the intellect makes a strong epistemological claim: just as sight, given a healthy organ, correctly perceives the colors of things in its visual field, so the FOI apprehends the quiddities, the real essences of sensible things. This stronger reading of veracity is well-supported by *lectio* 11 of *In De Anima,* which contains one of Aquinas's most extensive discussion of the veracity of the intellect.

There Aquinas, in commenting on Aristotle, contrasts the first and second operations of the intellect. The second operation is expressed in a complex judgment that is either true or false: "[Aristotle] says that the statement by which the intellect says 'something about something', as happens in affirmation, is always either true or false" (*In de anima,* III.xi.760). As said above, such a judgment involves both predicative *compositio,* the joining of discrete ideas in thought, and some assent; that is, the subject either takes an affirmative attitude in judgmental *compositio,* asserting that the mental synthesis exists in reality, or takes a negative attitude in *divisio,* denying such a correspondence. Such mental states are said to be true or false with respect to the *content* of the mental attitude; that is, they are veracious or fallacious with respect to a mental predication understood (according to Thomas) as corresponding, or not corresponding, to what is.

As we have seen, truth and falsity cannot enter into the FOI as it does in the SOI, for in the former there is as yet no judgment; in the FOI there is no mental attitude of judgmental *compositio* or *divisio,* which refers one's understanding to the individual things and their forms that are the proper objects of the intellect. Still, although in the FOI the subject does not refer her understanding to the objects of understanding, her intellect in understanding can be said to correspond with such objects. As Thomas writes of the FOI:

> But nevertheless, although this non-complex intelligibility itself is neither true nor false, nevertheless the intellect in understanding this [intelligibility] itself is true insofar as it is assimilated to the thing understood. And thus [Aristotle] adds that the intellect—"which is of this very 'what it is' [*quid est*] insofar as this is of something the what it is to be [*quod aliquid erat esse*]" (that is, insofar as it understands what a thing is [*quid est res*])—it [i.e., the intellect] is always true, and not insofar as it apprehends something about something (*In de anima,* III.xi.760).

The thing understood here, I submit, is the essence of that which has been perceived by the senses. The intellect in understanding is assimilated (*adaequatur*) to it insofar as its idea corresponds to this form, insofar as the essence is grasped in the idea. As Lonergan tells us, the term '*quod quid erat esse*' is generally used by Thomas "with special reference to the ground of essential definition, namely, the formal cause, so that at times it almost is, or simply is, the formal cause" (*op. cit.,* p. 23). The formal cause as the ground of essential definition is not in the intellect, but in the individuals with the form in question. The FOI is true, then, in that its ideas correspond to the essences of things. Let us say that when the intellect is true with respect to a certain idea in this sense it is ***directly true.***

As I am interpreting Aquinas, then, he is claiming that the FOI is always veracious in the sense that it is always directly true with respect to its ideas. That is, he is making the strong epistemological claim that the ideas of natural, essential kinds formed spontaneously in the FOI invariably correspond to the essences of things with whose ***phantasmata*** the intellect is presented. "Hence the intellect is never deceived in thinking what human being is (***quid est homo***)," as Aquinas writes (***In de anima,*** III.xi.761). Of course, in the prejudgmental stage of the FOI, the intellect does not understand or affirm that its apprehensions correspond to the essences of things; as Aquinas writes, "the intellect [is true] in thinking what something is: but not because it thinks or affirms the truth" (ST, 1.16.2c). That is, although the intellect does not understand its own understanding to be true, nevertheless, from the third-person point of view, the intellect in apprehending a quiddity can be said to have a true understanding of the essence.

IV. AN OBJECTION AND A RESPONSE

There is, however, an obvious objection to my stronger reading. Aquinas himself recognizes and discusses cases of imperfect and incorrect apprehensions of natural essences: he mentions our ignorance of the respective essences of flies,[11] bees (***ibid.***), and fire (ST, I.29. ad 3); he says that, because essences are often unknown to us, we must use accidents in place of substantial differences in our definitions (ST, I.77.1 ad 7); and, when our grasp of essences is deficient, we must use fallible discursive reason in coming to a full grasp of the essence.[12] It seems that, contrary to my claim about the veracity of the intellect, Aquinas thinks that our grasp of quiddities is liable to error.

Although I have so far spoken of the "stronger reading" as a single set of interpretive claims, it is necessary now to distinguish various versions of it. The passages just cited do undermine what we might call the ***strongest version of the stronger reading,*** which claims that the FOI fully and perfectly intuits the essences of all natural, material things that are perceived by the senses. Such a reading is both highly implausible in itself and at odds with Aquinas's texts. But then what sense can we make of a claims about the intellect's veracity?

---

[11] ***In symbolum apostolorum, prologus*** 864 in *Opuscula theologica,* II.
[12] ***Lectura super johannem,*** I.i.

Since a complete response to this question would require more space than is here available, I shall only sketch an outline of my solution. Essentially, I want to claim that, while the intellect is veracious in that it is competent to distinguish essences and classify individuals according to their essences—"to cut the world at its joints," if you will, it is often, at least initially, unable to specify the whole of the essence, or even the parts of the definition. On this interpretation, reasoning about ideas formed in the FOI is not employed somehow to establish the correspondence between these ideas and real essences, but to move from an imperfect to a full grasp of the essence. This interpretation we can still label the *strong version of the stronger reading* of the veracity doctrine, or just the *strong reading*.

Aquinas tells us that our intellect moves from an incomplete to a complete actuality, from a more universal, vague grasp of an essence to a less universal, definite grasp (ST, I.85.3c). Before we arrive at the complete actuality of "perfect scientific knowledge," there is often "imperfect scientific knowledge" in which "things are known indistinctly and with a certain amount of confusion" (ST, I.85.3c). Aquinas explains:

> It is manifest that knowing something comprised of many, yet without having a proper awareness of each of those of which it is comprised, is knowing something with a certain confusion. And one can know not only the whole universal (in which parts are contained potentially) but also the integral whole in this way: for in each case the whole can be known with a certain confusion, that is, without knowing the parts distinctly (ST, I.85.3c).

Thus, our initial grasp of an essence is sufficient to have an idea, to identify individuals of that essence and make competent judgments about them. And, indeed, such judgments issue in *scientia imperfecta,* imperfect scientific knowledge. Yet, one is unable to state what all the essential properties are.

In his commentary on the *Posterior Analytics,* Aquinas states that in imperfect scientific knowledge, although the whole of the essence is not apprehended, something of it is, as, for example, "if we grasp that human being is on the basis of the fact that human being is rational, while not yet knowing the other attributes which complete the essence of human being."[18] When one has a grasp of some part of the essence in this way, one grasps the essence partially and is in potency to a full apprehension.

---

[18] *In Libros Posteriorum Analyticorum,* III.xi.475.

It is important for us to appreciate, however, what the imperfect idea is *not*. It is not, for some essential attribute, the idea of whatever has that attribute. In the case of thunder, for instance, the claim is *not:*

(1) The definition of 'thunder' is "any noise in the clouds."

This would not be an idea of *thunder* at all, for, as Thomas says, "not every noise in the clouds is thunder" (*ibid.*, 475). If the cognitive grasp in question is to be the idea of thunder, then it must be an idea of the property *being a noise in the clouds* as *part of* a yet-to-be-discovered essence. This would involve a presupposition that there is some essence yet to be discovered, and some way in which that essence is signified. The idea must be that expressed by:

(2) The definition of 'thunder' is "a *certain sort* of noise in the clouds" (quaedam sonam in nubibus) (*ibid.*).

I am taking the 'certain sort of' as a reference to the as yet not fully known essence. The reference would seem to be in virtue of the instances of the properties that are taken as part of the essence; the phrase in question refers to their underlying cause. As he says in an analogous case, when speaking of using *accidentia propria* in place of unknown essential properties, the *accidentia propria* are used in the definition as "effects of the substantial forms, and as making these forms known (*manifestant eas*)" (ST, I.29.1 ad 3). That is, they are used to identify demonstrably the underlying essence that is to be discovered.

As the reader will recognize, I am attributing to Aquinas an *externalism for concepts,* the view, current in philosophical literature, that the individuation of at least some of our concepts depends not only on what is "in our mind" (as we have access to this through introspection), but also depends on the environment.[14] Returning now to the objection raised above, we can see that our initial grasp of quiddities may fall far short of full understanding—one may only apprehend part of the essence and remain ignorant of many or most essential properties. When this occurs, one's understanding may be expressed in a sort of nominal definition in which proper accidents stand in place of essential differences, as signifying the essence that is their cause. When we have such an imperfect grasp, we must employ fallible reasoning to come to a full grasp of the essence. In spite of such imperfect apprehensions, however, Thomas may still claim that our ideas, our *rationes,* do, in fact, correspond to the essences causally responsible for the *phantasmata* considered by the intellect, and, so, do correspond to the essences of things in the world. This amounts to the claim that the natural, essential kind classifications that the intellect spontaneously forms do, in fact, correspond to the real essences of what one perceives, and thus we need not expect a reworking of our conceptual structure in the light of further inquiry. Our intellects in their first operation do "cut the world at its joints," even though we may have much to learn about the essences so classified.

---

[14] A good example of a recent exposition of this sort of externalism can be found in Colin McGinn, "The Structure of Content," in *Thought and Object,* A. Woodfield, ed. (New York: Oxford, 1982)

V. CONCLUSION

If my strong reading is correct, then, although Aquinas may have thought that scientists of his time only imperfectly understood the respective essences of, for example, earth, air, fire, and water, he nevertheless believed they should be confident that these classifications corresponded to real essences. So understood, Aquinas's veracity doctrine is at odds with the subsequent development of the natural sciences: although water may still be thought to constitute a natural, essential kind, air and earth certainly do not. Indeed, significant progress was made in the natural sciences when the belief that air, for example, was an essential kind was eventually rejected and people came to understand it as any number of amalgams of elements that are gaseous at ordinary temperatures. And such progress would only have been possible after the rejection of Aquinas's doctrine of the FOI's veracity.

It may appear, then, that the result of my arguments in this paper have been to make Aquinas more distant from contemporary epistemological views and less interesting to contemporary philosophy. This would be so if figures in the history of philosophy interested us only to the extent that their views could be shown to be quite similar to our own best thoughts on a given matter—and, no doubt, many do have this attitude toward the history of philosophy. I believe, however, that it is the *differences* between ourselves and past philosophers which are as illuminating as the similarities, and that great philosophers of the past—of which Aquinas is certainly one—become really interesting when their philosophical distance from ourselves is recognized and acknowledged. I have argued in this paper for the recognition of one very important difference.

JOHN JENKINS

University of Notre Dame

EQUALITIES*

Feminist theories abound in most academic fields and yet their proponents often sense that the fields proceed without changing much in response to feminist arguments and critiques. I seek neither to verify this sense of marginal impact nor to criticize any given field for rendering feminist analysis marginal. But I do wish to note a conundrum. As long as feminist theories focus upon women and gender concerns, they run the risk of seeming marginal within fields of study that have neglected precisely those concerns. Yet to depart from a focus upon women and gender may seem to capitulate to the very definition of priorities that feminist analysis seek to challenge. Mindful of this conundrum, I propose to explore whether feminist analyses can illuminate issues that do not (or do not obviously) involve women or gender. Using recent Supreme Court decisions, I shall consider an aspect of the topic of equality, and specifically the legal doctrine, equal protection of the law.

The Supreme Court of the United States recently decided a case that asked whether equal-protection guarantees were violated when a prosecutor used his power to remove a limited number of jurors without stating a cause. In *Hernandez v. New York,*[1] the prosecution tried a case against a Latino criminal defendant and used the power to remove jurors to eliminate all potential jurors who were Latino. The prosecutor defended his actions by noting that two of the excluded jurors had family members who had been convicted of crimes and two others failed adequately to assure the prosecutor that they could defer to the official English translation of any Spanish-language testimony. Because prosecutors' "peremptory challenges" removed from the jury all Latinos who could have served on it, this is a promising and intriguing case for exploring the scope of equal-protection analysis.

I. *HERNANDEZ V NEW YORK* AND EQUALITY CONSTRAINTS ON PEREMPTORY CHALLENGES

The meanings and contours of equality are especially intriguing in the context of jury composition because two kinds of equality claims are involved: the claims of the litigants and the claims of potential

---

*To be read at an APA symposium on Feminist Critiques of Equality, December 30. Eva Kittay will comment. See this JOURNAL, this issue, 645–6, for her contribution. Thanks to Avi Soifer and Elizabeth V Spelman.

[1] *Hernandez v. New York,* 59 USLW 4501, 4502–3 (1981) (plurality opinion of Kennedy).

0022-362X/91/8811/633–44

jurors. Thus, in *Hernandez v. New York*, the criminal defendant could argue that the exclusion of Latino jurors violated both his right to equal protection of the laws (*ibid.*, 4502), and the rights of the potential jurors to have an equal opportunity to serve on a jury.[2]

This doubled quality of equality analysis winds through the history of challenges to jury-selection rules and practices. For over a century, landmark civil-rights litigation sought to eliminate race prejudice within the jury-selection process.[3] Historically, the discretionary actions of private attorneys in civil cases who exercise the privilege of the peremptory challenge to excuse jurors have been exempted from judicial rules about just causes for jury challenges. Yet the Supreme Court recently has ruled that even peremptory challenges by private attorneys are subject to constitutional bans against discrimination. Of no avail now are counterarguments that such challenges are purely private and should be immune from constitutional prohibition against governmental discrimination.[4]

Efforts to include gender within those bans against illicit discrimination have a convoluted history. In cases challenging the exclusion of women from juries through the 1960s, advocates confronted a dilemma posed by the then-prevailing conceptions of equality. Having to argue that women were like men in the respects relevant to jury service, the advocates met the response: If women are like men, why does the exclusion of women from the jury harm a female defendant?[5] Needing to maintain that women, just like men, would be competent jurors seemed to cut against claims that the exclusion of women deprived a female defendant of a quality of representation and understanding that only the female portion of the community could provide.[6] Ultimately, the Supreme Court interpreted the equal-protection guarantees to include women for the purposes of jury service. Interestingly, it did so in a case in which the defendant objecting to the exclusion of women was a man.[7]

---

[2] See *Edmonson v. Leesville Concrete Co.*, 1991 U.S. LEXIS 3021 (June 3, 1991) *9 (well-established rules of third-party standing allow a defendant to raise equal-protection claims of excluded jurors).

[3] See *Batson v. Kentucky*, 476 U.S. 79 (1986); *Swain v. Alabama*, 380 U.S. 202 (1965); *Strauder v. West Virginia*, 100 U.S. 303 (1880).

[4] See *Edmonson v. Leesville Concrete Co.*, 1991 U.S. LEXIS 3021 (June 3, 1991).

[5] Historian Linda Kerber explored this tension in her presentation about Hoyt v. Florida at the Harvard Law School, February, 1991.

[6] Cf. Susan Keating Glaspell, "A Jury of Her Peers," in *The Best Short Stories of 1917*, Edward O'Brien, ed. (1918), pp. 256-82 (1917 short story exploring possibility that women perceive and judge differently from men and nonetheless deserve participation in the system of justice).

[7] *Taylor v. Louisiana*, 419 U.S. 522 (1975).

The case I propose to examine, *Hernandez v. New York,* asks in part whether the equal-protection guarantees extend to ethnicity and communities defined by language as well as by race and gender distinctions. The struggle over language differences has produced important Supreme Court decisions in contexts outside the jury-selection problem.[8] It has also generated a political movement seeking to establish English as the official and preferred language.[9] The status of language differences in the jurisprudence of equality thus is a fresh and controverted issue presented by *Hernandez.*

Hernandez deeply split the Supreme Court.[10] The divisions among the justices reflected the prevailing disputes over equality jurisprudence but also indicated the assumptions shared across these different positions.

II PREVAILING DEBATE

None of the justices in the *Hernandez* case disputed prior rulings that demanded intentional discrimination to establish a violation of the equal-protection clause. Nonetheless, they differed over both the technical allocation of burdens of proof on this issue and the inferences to be drawn from the facts of this particular case.

Justice O'Connor, joined by Justice Scalia, argued that the case clearly lacked the requisite demonstration of intentional exclusion by race. Even if the peremptory challenges "may have acted like strikes based on race," she reasoned, "they were not based on race. No matter how closely tied or significantly correlated to race the explanation for a peremptory strike may be, the strike does not implicate the Equal Protection Clause unless it is based on race."[11] This rationale illustrates the quality of legal reasoning that separates that which is inextricably connected.[12] Another demonstration of this strange feat of separating the apparently inseparable appears in

---

[8] *Meyer v. Nebraska,* 262 U.S. 390 (1923) (constitutional rejection of statutory prohibition against public school instruction in German); *Lau v. Nichols,* 414 U.S. 563 (1974) (statutory guarantee of equality in schooling interpreted to call for accommodation for students lacking proficiency in English).

[9] See Thomas Weyr, "Hispanic U S.A.: Breaking the Melting Pot" 223-5 (1988) (describing English-only initiatives).

[10] Four members of the Court signed one opinion, authored by Justice Kennedy; two members signed another authored by Justice O'Connor; Justice Stevens wrote a dissent jointed by Justice Marshall, and Justice Blackmun dissented separately although indicated agreement with the essential elements of part of Justice Stevens' dissenting opinion.

[11] 59 USLW 4501, 4507 (O'Connor, J., concurring in the judgment).

[12] Thomas Reed Powell put it this way: "If you think you can think about a thing, inextricably attached to something else, without thinking of the thing it is attached to, then you have a legal mind." See John Gardner and Francesca Gardner Reese, *Know or Listen to Those Who Know: A Book of Quotations* (New York: Norton, 1975).

Justice O'Connor's claim that the equal-protection precedent in the jury-selection area "does not require that the justification [for a peremptory challenge] be unrelated to race. [It] requires only that the prosecutor's reason for striking a juror not be the juror's race."[13] She reasoned that, since the word 'race' was not offered by the prosecutor, race was not involved, even though the words offered by the prosecutor closely related to race.[14]

The four members of the court who signed the plurality opinion written by Justice Kennedy similarly but more elaborately reasoned that the prosecutor offered an explanation for the exclusions sufficiently unrelated to race. Thus, the defendant failed to establish that the prosecutor had engaged in intentional discrimination.[15] Three elements in this analysis are noteworthy.

First, the plurality deferred to the prosecutor's surface explanation and demanded no deeper inquiry; Justice Kennedy reasoned that "[u]nless a discriminatory intent is inherent in the prosecutor's explanation, the reason offered will be deemed race neutral" (*ibid.*). Second, the plurality rejected the defendant's claim that a close correlation between Spanish proficiency and Latino identity would be sufficient to treat exclusion of Spanish-proficient jurors as exclusions of Latinos, even though the plurality acknowledged that the exclusion of Spanish-proficient jurors had the effect of virtually excluding all Latinos (*ibid.*, 4503).[16] The disproportionate effect would be relevant only to the trial court's assessment of whether the prosecutor's rationale was merely a pretext for racial discrimination (*ibid.*, 4506). Third, while conceding that the defendant, by showing that all Latinos had been excluded from the jury, had carried its burden of establishing a "prima facie" case of an equal-protection violation, the plurality concluded that the defendant had not done enough to show that the prosecutor intended exclusions based on race.

The dissenters disagreed precisely with this third point. Justice Stevens reasoned that, once the defendant established the prima facie case of a violation, the requisite proof of intentional discrimination was present unless the prosecutor's defense specifically rebutted it. Moreover, the dissenters maintained that the prosecutor's own statements need not provide evidence of intentional discrimination (*ibid.*, 4509; Stevens dissenting). An indication of the subjective state of mind of the government actor would not be necessary, reasoned Justice Stevens, where other objective evidence sufficiently established intentional discrimination. Such objective evidence was present, he reasoned, because exclusion of Spanish-speaking persons would inevitably result in disproportionate disqualification of Latinos (*ibid.*). Finally, the dissenters found the prosecutor's explanation inadequate because less drastic measures could have been used to ensure deference to the official English translation (*ibid.*).[17]

---

[13] 59 USLW at 4507-8.

[14] The significance of Justice O'Connor's position may be greatest for the issue of what can count as evidence of intentional discrimination, and which party has the burden of proffering evidence on that issue—the complainant or the prosecutor, who must defend against a prima facie case of discrimination.

[15] 59 USLW at 4503.

[16] See also *ibid.*, 4505: "unless the government actor adopted a criterion with the intent of causing the impact asserted, that impact itself does not violate the principle of race-neutrality."

III. THE INADEQUACY OF PREVAILING EQUAL-PROTECTION ANALYSIS

At each of these points, the plurality and the dissenting opinions reflect the prevailing terms of equality debates in the federal courts. First, they differ over the significance of intentional discrimination compared with disproportionate results based on race or ethnicity. In their very clash of viewpoints about this debate, the opinions frame its centrality. The plurality assumed a sharp line distinguishing intent and effect or impact, while the dissent maintained their close relationship. In addition, the plurality demanded a "smoking gun" or direct evidence of subjective intent to discriminate on the basis of race or ethnicity, while the dissent was ready to infer sufficient intent from the results and the inadequacy of the prosecutor's defense (*ibid.*).[18]

As this disagreement suggests, however, establishing a distinction between intent and disproportionate impact is neither precise nor clear. It is a thin reed upon which to rest important assessments. Moreover, the very terms of the debate, as this summary may suggest, indicates the abstract quality of equality discussion currently underway in judicial interpretations of the equal-protection clause. Perhaps this is not quite the "emptiness" of equality analysis articulated by some observers,[19] but it does seem oddly both mechanical in form and indeterminate in application.

The plurality does render equality empty as a guarantee when it disconnects Spanish proficiency and race or ethnicity, and when it disconnects the group trait of Spanish-language proficiency from the individual responses of the potential jurors struck in this case. The plurality emphasized that these Latino potential jurors raised doubts for the prosecutor when they hesitated before answering that they would try to defer to the official English translation of Spanish testimony (*op. cit.*, 4504). The plurality ducked the defendant's response that this was an honest bilingual person's response. Even though the prosecutor's practice would have a disparate impact in excluding Latinos, according to the plurality, it did not use an explicit factor to exclude Latinos (*op. cit.*, 4504). The plurality concludes by rejecting the equal-protection challenge. Thus, the plurality:

---

[17] Justice Stevens also reasoned that the prosecutor could have argued for removing jurors "for cause," if he really believed that they could not defer to the official English translation. His failure to do so suggested some other concern that would not have satisfied the "cause" requirement.

[18] See Soifer, "Complacency and Constitutional Law," *Ohio State Law Journal*, XLII (1981), p. 383.

[19] See Peter Westen, "The Empty Idea of Equality," *Harvard Law Review*, XCV (1982), p. 537 (asking whether likes are treated alike lacks substance).

(1) argues for deference to any explanation offered for the discretionary use of a peremptory challenge that does not use race alone;
(2) refuses to view the prima facie case as establishing intentional discrimination sufficiently to withstand the prosecutor's defense;
(3) emphasizes the fact that some Latinos would indicate no reason to doubt their ability to defer to the English translation and some non-Latinos would give such doubt. Excluding those whose answers give rise to such doubt does not amount to an exclusion of Latinos (*op. cit.*, 4504), even though language and identity are closely linked

Behind the result and the doctrinal arguments lie assumptions that are contestable. Moreover, once elevated for viewing, those assumptions help to expose tensions and ambivalences within the justices' own analyses. Here, I maintain, feminist methods are quite useful.

IV FEMINIST INQUIRIES

Works by feminists identify assumptions underlying legal doctrines and theoretical debates in an effort to explore the room available for social transformation.[20] Concerned with the social structures that produce and maintain patterns of subjugation, feminist theorists emphasize the impact of power and position on knowledge and

[20] Similar work is now underway in law by "critical race theorists." See Mari Matsuda, "Voices of America: Accent, Antidiscrimination Law, and a Jurisprudence for the Last Reconstruction," *Yale Law Journal*, C (1991), pp. 1329, 1331 n.7. " 'Critical Race Theory,' as I use it, refers to the work of progressive legal scholars of color who are attempting to develop a jurisprudence that accounts for the role of racism in American law and that works toward the elimination of racism as part of a larger goal of eliminating all forms of subordination." See also Kimberlé Crenshaw, "A Black Feminism Critique of Antidiscrimination Law and Politics, in *The Politics of Law*, David Kairys, ed. (New York: Pantheon, 1990; 2nd ed.), pp. 195, 214 n.7: "Critical race theory goes beyond liberal understandings of race and racism by exploring those of its manifestations that support patriarchy, heterosexism, and class stratification. The normative stance of critical race theory is that massive social transformation is a necessary precondition of racial justice."

the significance of social and political context to any particular legal question. Exposing tensions present within legal texts can expose for critique the political choices surrounding those texts while stripping the appearance of inevitability or immutability the texts may exhibit.

A. *Exposing Assumptions.* Accordingly, feminist theorists stress and challenge three assumptions that usually remain unstated in analyses of equality and discrimination. The first is that the perspective of the excluded or subjugated person or group is irrelevant or untrustworthy in evaluating claims of discrimination: the perspective presumed relevant is that of those with the power to structure social institutions and to rule on charges of discrimination. The second assumption is that the equality inquiry does and should use as the norm for whether likes are treated alike those who have been privileged in the past: generally, white, male, Christian, English-speaking, able-bodied persons. The third assumption is that the status quo in social and economic institutions is sufficiently fair and uncoerced, and resists change.

In making these assumptions explicit, feminists suggest that the assumptions are themselves contestable and that alternative starting points should be used. Thus, they—we—suggest that equality questions and discrimination challenges should be evaluated at least in part from the perspective of excluded or devalued persons and groups.[21] We urge reconstructing the norm used to compare likes and unlikes in equality inquiries so that the full variety of human beings is included in the norm.

Rather than assume that a woman, for example, is sufficiently different to justify different treatment in employment, feminist critics challenge the presumed norm of a man used to define the woman as different. Our hope is that a reconstructed norm that includes both men and women will facilitate changes in workplaces and other practices that have treated women differently from men.[22] And rather than view the status quo as good, uncoerced, and in need of no major changes, the critics tend to view social and economic arrangements as vulnerable to criticism and reform.

As applied to equality, questions about prevailing assumptions suggest not resolutions so much as reasons to shift the debates to different territories. Thus, the contest over defining discrimination in terms of intent rather than impact seems a limited way of casting the contrast between the perspective of those with power to perpetrate or perpetuate discrimination and those victimized by it. An employer or government official may intend no discrimination or may be unconscious of it, yet nonetheless cause results that seriously injure or impair people in historically disfavored groups. Rather than debate the nature of the actor's intent, the critics urge that we acknowledge contrasting perceptions of what happened which expressed different people's positions in society. We seek to expose the contrast between perceptions and thus the choice involved in selecting one over the other. This translates the contrast between intentional discrimination and disparate impact into an open question of whose understandings to credit.

---

[21] See, e.g., Alan Freeman, "Legitimizing Racial Discrimination through Antidiscrimination Law: A Critical Review of Supreme Court Doctrine," *Minnesota Law Review,* LXII (1978), pp. 1049, 1052-7 (comparing the perspectives of victims and perpetrators of discrimination).

[22] See, e.g., *California Federal Savings & Loan Association v. Guerra,* 107 S Ct. 693 (1987) (permitting maternity leave as a method for equalizing male and female employees' abilities to have jobs and have families).

By shifting the inquiry through the vocabulary of contrasting perspectives, contestable norms, and a questionable status quo, feminists introduce to equality inquiries the impact of power on perception and knowledge, the significance of social and political context surrounding any given question of discrimination, and the exploration of historic patterns of domination. Feminists introduce different questions, therefore, for assessing any asserted violation of the ideal of equality, and recast the debates in ways that expose to public view deeper and broader dimensions than usually permitted by technical and doctrinal debates over equality. The feminist analysis exposes to view assumptions that otherwise insulate prevailing equality analysis from its social and political context, from the impact of power and perspective on knowledge, and from the facts of historic subjugation and contemporary fears about those unlike the dominant white population.

B. *Assumptions in Hernandez.* The works of feminist theorists help to illuminate assumptions about whose perspective matters to equality analysis, about what norm is chosen to assess like and unlike treatment, and about what aspects of social and economic arrangements are treated as acceptable and thus in no need of renovation. Asking about these assumptions in the context of *Hernandez* helps to expose how all the justices share these presuppositions:

(1) that the normal juror would not know Spanish, and the official English translation is the evidence that should count in the jury's deliberations;[23]

---

[23] It is not clear but possible that the justices also believe that language proficiency is not like race and thus not subject to illicit discrimination—though the plurality acknowledges significance of language to group identity. *Op. cit.*, 4606.

(2) that a pretextual rationale can be distinguished from illicit discrimination;

(3) that people who do not speak Spanish can adequately, and fairly judge people who do;

(4) that language proficiency is separable from ethnic or racial identity so significantly that language can supply a permissible basis for distinguishing people (while race or ethnicity may not) or that it can meaningfully be used in a comparison between intent and impact;

(5) that there is no need to consider or evaluate the Latino community's perceptions of a jury excluding all Latinos and yet sitting in judgment on a case involving a Latino defendant and Latino victims.

The effect of these unstated assumptions[24] for the plurality opinion is most curious. The assumptions lead Justice Kennedy to accept the trial court's conclusion that, because Latino jurors may be sympathetic to both the Latino defendant and to the Latino victims and witnesses, it is not discriminatory to exclude Latino jurors; the potential sympathies wash out.[25] This view neglects the perspective of Latinos in the community who may view participation in the trial as more than an expression of sympathies that can be canceled out and instead as a matter of participation in governance and in judgment of the conflicts within their community.[26] In addition, this view may imply that only Latinos have sympathies in a case involving Latinos.

Further, the plurality opinion treats the special knowledge Latinos may have as irrelevant to or even obstructive of justice. The plurality cites a case preferred by the prosecution "which illustrates the sort of problems that may arise where a juror fails to accept the official translation of foreign-language testimony."[27] There, a juror asked the judge if it would be possible to ask the translator about the meaning of a particular term. The translator had interpreted it to mean a public bar while the juror thought it meant a restroom. The judge indicated that questions could be put only to the judge, not the interpreter; the interpreter nonetheless volunteered that the jurors "are not to listen to the Spanish but to the English. I am a certified court interpreter" (*ibid.*).[28] At this point, the transcript produced by the court reporter indicates that the juror called the translator an "idiot," but the juror later explained she had said, "it's an idiom." The juror was dismissed from the jury (*ibid.*). Offered as evidence of the sort of problems that may arise where a juror fails to accept the official translation of foreign-language testimony,[29] this episode may also indicate the sorts of problems that arise when the trial process fails to accommodate people who are bilingual. The juror's question was treated as an intrusion rather than as an effort to get at the truth; she indicated that the witness's testimony would make more sense if it referred to a restroom rather than a bar, although later testimony confirmed that no meetings in a restroom had occurred. The court interpreter reacted defensively. And the judge responded by banishing the inquiring juror from the trial.[30]

---

[24] The reasons why the assumptions at work in this case remain unexplored by the justices seem closely linked to the issue of discrimination itself. To expose the assumptions to view is to raise the possibility of real change. That can be frightening. Full acceptance of someone who seems different can be frightening. Looking at the world from the perspective of another, reconsidering the norms used in analysis, and/contemplating changing elements of the status quo are all steps that could involve radical shifts in the understandings and practices used to manage and govern the world. These alternatives may be especially frightening to Anglos in the United States when Latinos or other Hispanics are involved The sheer numbers of Hispanics is large and growing quickly; one observer notes that "Sometime early in the next century Hispanic numbers will overtake those of blacks." Weyr, 229; see also 220 (in 1985, census estimates of the Hispanic population reached 17 3 million, up from 14.6 in the 1980 census, without including immigrants without legal status). Many Anglos fear Hispanic immigrants because they seem not to assimilate and instead maintain their language, customs, and communities more strenuously than do immigrants from other cultures. Weyr, 229. Weyr's exploration of the challenge to the Catholic Church represented by Hispanics provides a dramatic illustration of the persistence and insistance about maintaining their distinct practices and beliefs. Weyr, 190-218.

[25] 59 USLW at 4503, 4506 (citing and deferring to the trial court's reasoning).

[26] Cf. Randall Kennedy, "*McCleskey v. Kemp:* Race, Capital Punishment, and the Supreme Court," *Harvard Law Review,* CI (1988), p. 1388.

This story contrasts sharply with another case in which a man got into a fight in a bar with another man and killed him. Both men were Mexican-Americans. The offender argued that his victim had given him "*el ojo,*" meaning, the eye. At that time, no Mexican-Americans were eligible to serve on the juries in Texas where the incident occurred. The defendant was convicted of murder. As one observer notes about the case:

> Anglos have a big thing about eye contact being something positive. You can take a man's measure by making eye contact - - - . Hell, in the Mexican community eye contact can kill you. It sends the other guy

---

[27] 59 USLW at 4504 n.3 (quoting from *United States v. Perez,* 658 F.2d 654 (CA9 1981))

[28] (quoting *United States v. Perez,* 658 F.2d 654, at 662).

[29] Foreign itself is an interesting word, since it uses English as the norm and all other languages are "foreign," even if they are the primary language of the witnesses, the defendant, and even some of the jurors.

[30] Query: Is language proficiency a more threatening kind of community knowledge than other kinds of knowledge that jurors are expected to bring to the deliberation process? See below (discussing fears of non-English speakers in the United States).

> a message that says what the hell are you lookin' at, and if you don't like it, do something about it. In a bar that can lead to a killing. But if you don't know that you can't relate to what it means. And unless jurors understand the difference between *el ojo* and eye contact, the defendant is not being tried by a jury of his peers.[31]

In 1954, the Supreme Court of the United States—composed by different justices than the current court—reversed the conviction on this kind of rationale (*ibid.*, 83).[32]

As the result in this second case suggests, the status quo can change. The courts sometimes have so ordered. Yet the very possibility of changing the ground rules to avoid the problem perceived by the prosecutor who struck the Spanish-speaking jurors in *Hernandez v. New York* remained somehow obscured by the doctrinal arguments preoccupying that court. If only the justices had discussed this issue—whether a change in the underlying rules governing jurors' roles and practices—they might have noticed that a clear majority of the justices did indeed favor a change. The four justices in the plurality and both Justices Stevens and Marshall in dissent—six justices out of the nine—considered ways to change the status quo of the jury to accommodate the bilingual jurors. The plurality and Justice Stevens, joined by Justice Marshall, suggested that jurors proficient in a language used by witnesses could be given an opportunity to indicate to the judge a problem with the translation.[33]

If the justices had treated the rules governing the bilingual juror as a central concern, they could have explored even more radical changes. The court could require some or all of the jurors in a case involving a defendant whose primary language is not English to share proficiency in the defendant's language. This possibility was

[31] Gilbert Pompa, quoted in Weyr, 83.

[32] See *Hernandez v. Texas*, 347 U.S. 475 (1954)

[33] 59 USLW at 4504, 4565 (Kennedy, J.); 4509 (Stevens dissenting): "For example, though petitioner did not suggest the alternative to the trial court here, Spanish-speaking jurors could be permitted to advise the judge in a discreet way of any concerns with the translation during the course of the trial." This suggestion implies that Spanish-speaking jurors would not have to defer to the English translation under all circumstances, as the prosecutor in *Hernandez* demanded. Justice Stevens noted that "bilingual jurors could have been instructed to bring to the attention of the judge any disagreements they might have with the translation so that any disputes could be resolved by the court." He continued in footnote that "An even more effective solution would be to employ a translator, who is the only person who hears the witness's words and who simultaneously translates them into English, thus permitting the jury to hear only the official translation" (n.2). While this indeed would solve the problem perceived by the prosecutor in this case, it neglects an additional problem potentially perceived by the Spanish-speaking defendant and witnesses: their chance to be fully heard would be curtailed, and the jury's role in evaluating credibility would also be hampered.

not even debated by the justices, if the written opinions are any indication of the scope of their debate.

Yet even without this more radical suggestion, the plurality opinion is not so distant from the concerns raised by a feminist analysis. That opinion offers the most explicit Supreme Court acknowledgment of a close relationship between language and identity and between language and ethnicity, sufficient at times to justify equal-protection scrutiny on the basis of language proficiency (*ibid.*, 4505, 4506–7).

In addition, at least on one occasion, the plurality looks to the perspective of the excluded Spanish-speaking juror and comments that it would be a "harsh paradox that one may become proficient enough in English to participate in trial," given requirements of English-language ability for federal jury service, "only to encounter disqualification because he knows a second language as well" (*ibid.*, 4506). This kind of insight could be an opening for justice rather the irony if treated as a central concern. Feminist analysis would support just that emphasis and just that movement to the center of what, and who, have been treated as marginal.

C.*Feminist Techniques, Human Consequences*. I have suggested here how feminist inquiries about assumptions prevailing in legal analysis of equality can shift the terms of analysis in a context involving Spanish-speaking individuals. This effort is not undertaken with a belief that all forms of discrimination or group-based disadvantage are the same. Instead, switching to feminist questions holds more promise for invigorating equality and ensuring its role as a continual prod to re-evaluate and re-examine who we each are and who we all can become.

MARTHA MINOW

Harvard Law School

IN WHOSE DIFFERENT VOICE?*

The presence of a feminist legal theorist allows us to reflect on shared assumptions in legal theory and philosophy which leave little space for feminist reflections: a universal model of the legal subject, and a concept of justice emphasizing universality, impartiality, and neutrality. Group interest is an unwelcome intrusion of difference into the ideal of equality. Partiality and perspective threaten to tear the benevolent blindfold off the figure of justice.

No one who pursues these concerns desires to undermine hard won assumptions of moral and political parity. Still, some insist that the liberal ideals of impartiality, neutrality, and a policy of treating likes alike are impotent in bringing about the egalitarian vision that justifies them. Martha Minow's has been a major voice insisting that equality will be formal or even empty, until perspective and difference are acknowledged and incorporated within the fabric of legal theory and practice.

As her title (*Equalities*) suggests, the question of equality fragments into one of equalities. Equality *of what?* Equality *for whom?* Equality *by what measure?* Minow asks us to consider the different perspectives that deem one situation equal to some and unequal to others, and asks us to consider to whom we ask to be equal.

Minow focuses on issues of language, ethnicity, and exclusion. Although *Hernandez v. New York* does not have an immediate impact on feminist concerns, it brings to bear questions of perspective and voice. These occupy the spotlight when we ask how those who vary from the norm can receive equal protection under the law without being cut to fit a procrustean bed. I intend to examine how the conclusions Minow sets forth can be used to resolve debates within feminist legal theory. Will the suggestions put forward here succeed in resolving for us what Minow has elsewhere called the "dilemmas of difference"?

The remainder of my remarks will be addressed to Minow's stated commitment to use feminist critiques to illuminate cases of exclusion not based on gender. This strategy, she suggests, will at once counter the dangers of the marginalization of feminism and contribute to the "engendering" of justice for all. Although this is persua-

---

* Abstract of a paper to be presented in an APA symposium on Feminist Critiques of Equality, December 30. Commenting on a paper by Martha Minow. See this JOURNAL, this issue, 633–44, for her contribution.

0022-362X/91/8811/645–6

sive, it is worth considering that the force of the feminist critiques resides in the pervasiveness and near universality of woman's subordinate status.

Feminist scholars today are engaged in a searching inquiry. Not since Descartes have so many assumptions about what it is to be an inquiring subject been questioned. Not since Marx have so many deep assumptions about political and social structures been subjected to scrutiny. In the process, feminists have had to look at the ways in which different forms of oppression interweave, undercutting the simple opposition of *us* and *them,* as factors that splinter the unity of the first-person plural, as well as the hegemony of the third-person plural.

I urge that we need to resist the idea that only where and when feminist work can be applied and generalized can it avoid marginalization. Oppression by gender, while historically and culturally and class specific in its manifestations, is as close to a universal social phenomenon as we have.

At the same time, recent feminist work has cast doubt on the validity of analogies or identifications of different forms of oppression, such as those often posited between race and gender. Without ignoring different forms of oppression, we need a clear focus on women's lives, else we miss the extent to which socially assigned roles (especially the role of caregiver) themselves make possible the social order valorized in liberal political and legal theory. To conceive of society as an association of equals ignores the inevitable dependency of some of its members, a dependency that makes both these individuals and their caregivers unable to compete on equal terms in the economic, political, and social domains. I shall argue that the nub of the problem of inclusion of women within the ideal of equality can best be seen by looking at women in their role as dependency worker.

EVA FEDER KITTAY

State University of New York/Stony Brook

CAN BIOLOGICAL TELEOLOGY BE NATURALIZED?*

Naturalism is the doctrine that everything real is at least in principle within the scope of a purely scientific account of the world. Although the naturalistic umbrella contains ample room for divergent views, it is widely agreed to exclude moral and, more generally, evaluative matters. A naturalistic theory must be "nonnormative," as Ruth Millikan[1] insists, erecting some form of a relatively rigid fact/value distinction and banishing "transcendent standards of value" (to use D. M. Armstrong's[2] words) from nature. Let ***narrow naturalism*** be this naturalism-cum-descriptivism. (Unqualified references to naturalism in what follows always apply to narrow naturalism.)

There is a growing consensus that biological teleology can be narrowly naturalized.[3] Naturalistic treatments of biological teleology tend to be modeled after either cybernetics or natural selection, and the majority conviction is that eventually some version of one of these approaches will secure a naturalistic treatment of biological teleology. This paper aims to undermine this consensus confidence by revealing an evaluative component of biological teleology.[4] Space limitations permit only a partial presentation of this argument. Elsewhere I have argued that the cybernetic approach to naturalizing biological teleology fails because of its inability to account for an evaluative component in teleology.[5] Here I shall sketch a parallel

---

* To be presented in an APA symposium, Can Biological Teleology Be Naturalized? December 30. Mohan Matthen will comment. See this JOURNAL, this issue, 656–7, for his contribution.

[1] *Language, Thought and Other Biological Categories* (Cambridge: MIT, 1984), p. 17.

[2] "Naturalism, Materialism, and First Philosophy," *Philosophia*, VIII (1978): 261–76, p. 263.

[3] For example, see Ernest Nagel, "Teleology Revisited," this JOURNAL, LXXIV, 5 (May 1977): 261–301, pp. 276f, 298ff; and Larry Wright, *Teleological Explanations* (Berkeley: California UP, 1976), pp. 26f, 56f.

[4] Naturalistic accounts of biological teleology have recently been used to save naturalistic treatments of mind. For an outline and references, see William Lycan, ed., *Mind and Cognition, A Reader* (Cambridge: Blackwell, 1990), pt II. Thus, the failure of naturalistic approaches to biological teleology would not only undo what was banked on as a naturalistic done deal; it would also block an attractive route to naturalism in the philosophy of mind.

[5] An extended discussion of the cybernetic approach to teleology appears in my "Goal-Directed Systems and the Good," forthcoming in *The Monist*. A vastly expanded treatment of these arguments is forthcoming in my "Naturalism and Teleology," in S. Wagner and S. Warner, eds., *Naturalism: A Critical Appraisal* (Notre Dame: University Press). A positive value-centered account of teleology is contained in my "Where's the Good in Teleology?" *Philosophy and Phenomenological Research*, forthcoming.

0022-362X/91/8811/647–55

argument concerning the attempt to naturalize teleology by appeal to natural selection.

I

One main naturalistic approach to teleology keys on certain kinds of etiologies. The idea of this approach is, roughly, that something happens in order to achieve a goal when (i) it tends to bring about that goal and (ii) it happens because of its tendency to bring about the goal.[6] Consider the heart pumping blood in order to circulate it. According to the etiological approach, this means that (i) the heart's pumping of the blood tends to circulate the blood and (ii) the heart pumps blood because doing so contributes to circulating it. Certainly, the heart's pumping does circulate the blood, but why does the heart pump blood? Well, the heart would not be pumping unless the creature possessing it were alive, and the creature would not be alive unless its blood were being circulated. If the heart were to stop circulating the blood, the creature would die and its heart would stop pumping. So, the heart pumps in part because doing so contributes to circulating the blood.

There is much to be said for this approach to teleology,[7] but as it stands it is not sufficient—something that is not teleological might nevertheless have an etiology like the heart's. Consider a stick floating down a stream which brushes against a rock and comes to be pinned there by the backwash it creates.[8] The stick is creating the backwash because of a number of factors, including the flow of the water, the shape and mass of the stick, etc., but part of the explanation of why it creates the backwash is that the stick is pinned in a certain way on the rock by the water. Why is it pinned in that way? The stick originally became pinned there accidentally, and it remained pinned there because that way of being pinned is self-perpetuating. Therefore, once pinned, part of the explanation for why the stick is creating the backwash is that the backwash keeps it pinned there and being pinned there causes the backwash. In this case, the stick meets the etiological conditions: creating the backwash tends to pin the stick on the rock and the stick creates the backwash because doing so contributes to pinning it. Clearly, however, the stick does not create the backwash *in order to* keep itself pinned on the rock. The stick's behavior has no teleological explanation.

---

[6] Wright (*op. cit.*) has offered the best-known version of the approach.

[7] In fact, the value-centered view that I defend in "Where's the Good in Teleology?" can be viewed as a modification of Wright's analysis to which a value condition has been added.

[8] This example (though not its diagnosis) is due to Robert Van Gulick.

Given that the heart and the stick involve a similar sort of etiology, why is only the heart teleological? There are two ways to revise the etiological approach to answer this question: to appeal to value or to appeal to natural selection. One clear difference between the two cases is that the heart's pumping is good for the creature, but the stick's backwash is not good for the stick (and it need not be good for anything whatsoever). Thus, one diagnosis of the stick in the stream counterexample is that teleology must involve an appropriate role for value. Teleology involves not merely effects that tend to bring about their own production, but good effects that tend to bring about their own production.[9] Those sympathetic with the etiological approach might offer an alternative diagnosis, in terms of natural selection. Roughly, the idea would be that a feature is for the sake of any effect by virtue of which it is naturally selected, i.e., by virtue of which natural selection favors it.[10] This would handle the stick example because natural selection does not explain why the stick is stuck on the rock. Sticks floating in streams do not reproduce their kind. Being (able to be) stuck on a rock is not a trait inherited from parent sticks, nor is it a trait that affects a stick's "fitness." So, the stick in the stream is not a counterexample to a version of the etiological approach which explicitly invokes natural selection. Perhaps the value diagnosis is wrong and teleology can be naturalized by appeal to natural selection.

II

The value diagnosis and the natural selection diagnosis each handle the stick in the stream. We need a case that drives a wedge between them. But it is difficult to find a value-free case of natural selection. The familiar biological examples of natural selection will not do, since the traits by virtue of which natural selection favors an organism typically benefit the organism. For example, once the forests around industrial centers in England became besooted by industrial pollution, the whitish forms of the peppered moth *Biston betularia* became much easier for predatory birds to find. In those environments, industrial melanism (black coloration) benefited the peppered moth by camouflaging them.[11] What we need is a case of natural selection in which value clearly plays no role whatsoever.

---

[9] There are different grades of teleology corresponding to different etiological roles for value; see my "Where's the Good in Teleology?"

[10] Wright points out that his analysis can be understood in terms of natural selection (*op. cit.*, pp. 69f). Others, inspired by Wright, explicitly refer to natural selection in accounts of natural teleology; e.g., see Fred Dretske, "Misrepresentation," in R. Bogdan, ed., *Belief: Form, Content, and Function* (New York: Oxford, 1986), p. 25; D Papineau, *Reality and Representation* (Cambridge: Blackwell, 1987), p. 65; and Mohan Matthen, "Biological Functions and Perceptual Content," this JOURNAL, LXXXV, 1 (January 1988): 5–27, pp. 14, 17.

To find such a case, it is helpful first to focus on the idea of a value-free situation. It seems that, everything else being equal, surviving is good for an organism. An organism seems worse off if it is barely eking out an existence; it seems better off to the extent that it is flourishing. One way to guarantee a value-free situation would be to have one free from all forms of life. So, my strategy in the following pages will be, first, to describe a lifeless planet, devoid of teleology, and also devoid of natural selection. Then, I shall add to the planet what is required for natural selection to occur among certain inorganic minerals. But at no stage will the planet contain any form of life.

Consider a planet containing land and sea masses surrounded by an atmosphere. In addition to certain atmospheric gases (oxygen, hydrogen), the planet contains various relatively simple mineral compounds, such as those found in rocks and sand (silicate, aluminium, etc.). The stuff on this planet is subject to the familiar orchestra of physicochemical processes, from Born repulsion to covalent bonding. Sand and dust are blown about, rains fall, rocks weather, particles become suspended in water and then are sedimented, minerals dissolve and crystalize out of solution, chemical bonds form and break. Various physicochemical equilibria are disturbed and restored; various cyclical physical processes occur (a water cycle, a tidal cycle, daily and seasonal temperature cycles, etc.). But the planet is lifeless. It contains absolutely no forms of life, not even any organic molecules. Not only is this dead planet value-free, I take it to be uncontroversial that there is nothing teleological on it. Nothing has any purpose, nothing is for the sake of anything. Everything that happens on the dead planet is a paradigm of an ordinary efficient causal process. The dead planet would be thought to involve teleology only by someone, like Teilhard de Chardin perhaps, who finds teleology in virtually every merely physical process—in water flowing downhill or in the cycle of the tides. In addition to lacking value and teleology, the dead planet also lacks natural selection. Specifically, it lacks each of the following four necessary and jointly sufficient conditions for natural selection: a population of *reproducing entities*, the members of which possess *randomly varying* traits which can be *inherited* and which are *adaptive* (in the sense that they can causally influence the extent to which entities with the traits proliferate).[12]

---

[11] A nice review of the empirical evidence for this famous example of natural selection in wild populations can be found in J. Maynard Smith, *The Theory of Evolution*, 3rd ed. (New York: Penguin, 1975), pp. 150–4.

It turns out that all four prerequisites for natural selection can be added to the dead planet. What is rather surprising is that, in the process, the planet remains entirely devoid of life. Thus, natural selection can take place in an entirely inorganic setting. This is possible because of the nature of crystals, specifically, the microscopic crystallites of inorganic materials out of which clays and muds are composed.[13] When we think of crystals, we tend to think of something on the order of fingernail- or fist-sized chunks of quartz, amethyst, mica, or possibly diamond. The last thing that comes to mind is clay. But if you were to magnify particles of clay thousands of times with a scanning electron microscope, you would see an amazing diversity of tiny crystallites. Populations of clay crystals have each of the four features necessary for natural selection. In fact, natural selection is occurring in populations of clay crystals all over the earth right now. Consider each prerequisite for natural selection in turn.

Reproduction. Crystals are simply layers of orderly arrays of molecules, which grow as new layers are added on top of existing layers. Clay crystals are formed basically from silicic acid (SiO_4) and hydrated cations (positively charged ions). These crystals can form in the presence of a supersaturated silica solution if crystal "seeds" (small pieces of crystal) are present. A new layer grows as silica atoms from the solution become affixed to the surface of the seed. When a crystal becomes big enough it cleaves and pieces break off, becoming seeds for new crystals. In this fashion, crystals reproduce. Clay crystals tend to break up easily during their formation, as a result of cyclical and sporadic variations in ambient conditions, similar to the way in which the layers of a mica crystal (not a clay, but a relatively easily studied prototype for clays) can separate because of changes in external conditions, such as the presence of certain ions. Just as a population of reproducing biological organisms is the setting for biological evolution, a population of reproducing crystals can be the setting for crystal evolution.

---

[12] Richard Lewontin, "The Units of Selection," *Annual Review of Ecology and Systematics*, I (1970): 1–18, p 1; A. G. Cairns-Smith, *Seven Clues to the Origin of Life* (New York: Cambridge, 1985), p. 2; R. Dawkins, *The Selfish Gene* (New York: Oxford, 1976), pp. 150ff.

[13] My information about crystals in general, and clay crystallites in particular, is due to Cairns-Smith: "The Origin of Life and the Nature of the Primitive Gene," *Journal of Theoretical Biology*, X (1965): 53–88; *Genetic Takeover and the Mineral Origins of Life* (New York: Cambridge, 1982); and "The First Organisms," *Scientific American*, CCLII, 6 (1985): 90–100. I had long thought that the natural selection reading of the etiological approach was vulnerable to hypothetical counterexamples involving value-free cases of natural selection, and I speculated that crystals might be an appropriate setting for such hypothetical cases. Upon discovering Cairns-Smith's discussion of clay crystallites (written for an entirely different purpose), I was delighted to learn that the counterexamples I had imagined were not merely hypothetical but actual.

Variation. In an ideal world, all crystals of a given substance would be exactly alike in structure. Each part of an ideal crystal would have a molecular lattice structure that is just like that found in every other part. In our actual world, all crystals have flaws. These crystal defects are produced haphazardly, "randomly," just as genetic mutations are, and the potential variety of crystalline defects is enormous. These flaws can be a source of novel information in the evolution of crystal populations, just as genetic mutations are in organic evolution. Flaw-produced novelties can be expressed in a wide variety of "phenotypic" traits. For example, physical defects in crystals can affect a crystal's shape, its growth rate, and the conditions under which it cleaves. In addition, the chemical defects in microscopic crystallites can play a dominant part in determining their physicochemical properties, such as whether they remain dispersed, coalesce into loose clusters, or form a gel; even the density of the gel can be affected. Thus, crystalline defects cause the sort of random variation in phenotypic traits required by natural selection.

Heredity. When a new layer of a crystal grows, it copies the geometrical arrangement of atoms occurring in the layer below. For example, new layers in a graphite crystal copy the flat hexagonal pattern of the carbon atoms in the lower layers of graphite, while new layers in a diamond crystal copy the tetrahedral pattern of the carbon atoms in the lower layers of diamond. Crystalline defects can be copied in the same way. For example, as atoms in new layers align with the atoms below, any grain boundaries or dislocations tend to be copied and, thus, the defect propagates in subsequently added layers. When a crystal with such a defect cleaves and reproduces, the crystalline offspring will contain the same defect, and the defect will then propagate through the layers as these new crystals grow. When these new crystals themselves cleave, the defect will be passed on to their offspring. In this way, the defects and the phenotypic traits they cause can propagate from generation to generation. If further defects occur, these too can propagate through the crystal population. Many "mutations" later, a population of crystals can contain many different crystal "species," each with its own characteristic pattern of crystalline flaws and resulting phenotypic traits. These different species can provide the subpopulations among which nature selects.

Adaptivity. For natural selection to get a grip on the subpopulations, the inherited variations must have the ability to affect the crystal's survival and reproductive success—its fitness. There are many ways in which crystalline phenotypic traits might accomplish this. A crystal's shape, the rate at which it grows and the conditions under which it cleaves (reseeds) can all affect the rate at which that crystal "species" proliferates. A crystallite species' physicochemical properties, such as whether the particles remain dispersed, coalesce into loose clusters, or form a gel, can affect their dispersal tendencies. The density of such gels can affect the diffusion rate of the clay-forming solution and thus the rate at which they reproduce. These are just some of the ways in which crystal traits can bear on the relative fitness of different crystal "species." An example from Dawkins[14] nicely illustrates how a clay crystal's defect-driven traits can be adaptive.

> Whether or not a particular type of clay crystal is allowed to build up depends, among other things, upon the rate and pattern of flow of the stream [in which the chemical building blocks of crystals are dissolved]. But deposits of clay can also *influence* the flow of the stream. . . . [S]uppose that a variant of a clay improves its own chances of being deposited by damming up streams. This is an inadvertent consequence of the particular defect structure of the clay. In any stream in which this kind of clay exists, large, stagnant shallow pools form above dams, and the main flow of water is diverted into a new course. In these still pools, more of the same kind of clay is laid down. A succession of such shallow pools proliferates along the length of any stream that happens to be 'infected' by seeding crystals of this kind of clay. Now, because the main flow is diverted, during the dry season the shallow pools tend to dry up. The clay dries and cracks in the sun, and the top layers are blown off as dust. Each dust particle inherits the characteristic defect structure of the parent clay that did the damming, the structure that gave it its damming properties. . . . The dust spreads far and wide in the wind, and there is a good chance that some particles of it will happen to land in another stream, hitherto not 'infected' with the seeds of this kind of dam-making clay. Once infected by the right sort of dust, a new stream starts to grow crystals of dam-making clay, and the whole depositing, damming, drying, eroding cycle begins again (*ibid.*, p. 153).

Now that we appreciate the possibility (indeed, actuality) of natural selection among crystals, we have the ingredients of a value-free case of natural selection. Thus, we can finally assess the version of the etiological approach to teleology which appeals to natural selection, for we can now drive a wedge between the value diagnosis and the natural selection diagnosis of the stick in the stream counterexample.

---

[14] *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design* (New York: Norton, 1986).

We have seen that populations of clay crystals can have all the elements necessary for evolution by natural selection. Consider the crystalline defect (call it *D*) that is responsible for the damming action of Dawkins's clay. Notice that natural selection causes *D* to persist in the clay population because of its "advantageous" dam-making effects. In short, just as natural selection favors the black subpopulation of the peppered moth in those environments blackened by industrial pollution, natural selection favors the clay species containing *D* in the environment described by Dawkins. But, while we are inclined to say that the moth's dark coloring is to camouflage them from predatory birds, we would *not* want to say that, as a matter of natural teleology, the *purpose* of *D* is to make dams, that *D* is there *for the sake of* making dams or *in order to* make dams. Thus, when a process of natural selection causes a trait to persist in a population because of its effects, there is no guarantee that the trait will be for the sake of that effect. The trait might not be for the sake of anything. So, even though the dead planet has the physicochemical resources to support a full-blown process of natural selection, it is still not the kind of place where there is any teleology. Merely having a population of replicating entities will not produce teleology, not even if the entities have random variations that are heritable, not even if those variations can affect the degree to which those entities proliferate. What happens on the dead planet has no purpose and is for the sake of nothing; it is merely a physicochemical process. Therefore, contrary to those versions of the etiological approach to teleology which are couched in terms of natural selection, natural selection is not sufficient for teleology. A feature might not be for the sake of the effect by virtue of which it has been favored by natural selection; a feature that has been naturally selected might not be for the sake of anything.

To summarize: the presence of natural selection in populations of crystals allows us to check whether the etiological approach to teleology which appeals to natural selection captures a sufficient condition, for each condition in the approach can be satisfied by features of crystals. But, intuitively, features of crystals are not the sort of things that fall within the realm of teleology. Thus, we have a clear-cut counterexample to the natural selection version of the etiological approach to teleology.

III

How should we diagnose this counterexample to the natural selection treatment of teleology? We do think that the peppered moth's dark coloring is for the sake of protecting it from predators, that the heart pumps blood in order to circulate it—in general, that those traits in biological organisms favored by natural selection are clear-cut instances of natural teleology. This raises the question: What differentiates natural selection in crystal populations and natural selection in biological populations? If natural selection among crystals is insufficient for teleology, then why is there genuine teleology in the biological products of natural selection? Here the value diagnosis reasserts itself: value notions apply in the biological world but not on the dead planet. Dark coloring is good for peppered moths in industrial environments; but it is clearly a metaphor at best to say that mineral property *D* is "good" for Dawkins's dam-making crystals. Value notions apply to living things, even those which are not human. For horses, birds, bees, plants, even amoebas, the notion of being made better or worse off by a trait makes sense. On the other hand, these value notions do not make sense when applied to crystals, rocks, and other bits of inorganic matter, except perhaps metaphorically. So, as with the original stick in the stream counterexample, the value diagnosis explains away the crystal counterexample by pointing to the absence of an appropriate role for value.

The narrow naturalist might resist this value diagnosis. Although space limitations prevent any discussion, let me at least indicate two concerns. (i) The naturalist might bite the bullet and insist that value-free clay crystallites can be teleological, but this move leads to an implausible radical panteleologism. (ii) The naturalist might attempt to rule out crystal teleology by restricting consideration to natural selection in living populations, but it proves difficult to give a noncircular narrowly naturalistic account of life.[15]

These difficulties with naturalistic accounts of biological teleology arise specifically for the narrow naturalism that banishes "transcendent standards of value" from nature. But naturalists need not be narrow. A broader view of nature, perhaps roughly Aristotelian in outlook, could reckon objective standards of value as part of the natural order. According to this broader form of naturalism, which would contrast with supernaturalism and would reject the miraculous in nature, values would be real ineliminable natural properties, subject to broadly scientific investigation. Making sense of this broadly construed naturalism might enable the many attractions of a naturalistic treatment of biological teleology to be realized.

MARK BEDAU

Reed College

---

[15] Norman Packard and I discuss the connection between teleology, life, and evolution in "Measurement of Evolutionary Activity, Teleology, and Life," in C. Langton, C. Taylor, D. Farmer, and S. Rasmussen, eds., *Artificial Life II* (Redwood City, CA: Addison-Wesley, 1991).

NATURALISM AND TELEOLOGY*

ON the face of it, teleology is based on the idea that things in the natural order are constructed for a purpose, or subserve some good. Naturalism tries to eliminate precisely this sort of supposition from the physical and biological sciences. Thus, naturalists have always been embarrassed by the persistence with which biologists use ends. Hence the attempts to construct *validations* of teleology, to show that the factors cited in teleological explanations are in fact explanatory, but that, contrary to appearance, they presuppose nothing unacceptable to naturalism. Most such validations of teleology are in fact *reductions*—claims that each teleological predicate *Fxy* (for example, 'The function of *x* is *y*') is equivalent (logically, analytically, or materially) to some naturalistically acceptable predicate *Gxy*, where *Gxy* really does perform the explanatory work that *Fxy* is thought to perform.

Mark Bedau argues against the two leading types of reduction. One horn of his dilemma involves circularity. The reduction implies that crystals might have functions, he argues convincingly—but, he says, teleology cannot be applied to lifeless things because value cannot. But it is Bedau, not the naturalist, who is trying to advance the "value diagnosis" of teleology. The other horn queries the *emergence* of teleology: "Merely adding a reproducing population of mineral entities seems insufficient to create teleology." Thus, the view involves "a slippery slope with radical panteleologism . . . at the bottom." But this shows a misunderstanding of how naturalistic validations are constructed. There are, presumably, certain explanatory tasks that functions are supposed to perform—for example, the function of a trait is supposed to be an effect that is also the reason why the trait is a stable feature of things of a certain sort. If the reduction saves these explanations, it is acceptable. "Panteleologism" should not threaten the enterprise—not all stable traits are so because of their effects, and so not all have functions. And how does the question of emergence arise? It is a problem only if you think that functions are something strange and different. Naturalists do not.

But look at the matter in a different way. Suppose that Bedau's intuitions about functions are right. His counterexamples would

---

* Abstract of a paper to be read in an APA symposium, Can Biological Teleology Be Naturalized? December 30, commenting on a paper by Mark Bedau. See this JOURNAL, this issue, 647–55, for his contribution.

0022-362X/91/8811/656–7

then show that the intuitive notion of function cannot be *reduced* to something naturalistically acceptable. This does not imply that its biological applications cannot be *validated*. I favor the "etiological" approach to validation myself: 'The function of *x* is *y*' is genuinely explanatory only if *y* adds to the fitness of bearers of *x*. It may be that this approach lumps the functional claim together with certain other explanations that are not intuitively functional. What of it? The intuitions are remnants of the pre-Darwinian worldview; it may well be impossible to save all of them. Validating some explanatory uses of functions may serve the naturalist well enough.

Bedau's own account of teleology is nonvalidating. To the etiological analysis he adds a value condition. But if there is one thing that has become completely clear in recent years, it is that natural selection does not favor valuable traits. Indeed, it now appears that the idea that evolution constitutes progress was essentially a political ideology.[1] So to allege that functions are essentially valuable is to forego the possibility of a validating reduction of the concept, at least if the naturalist is right about the nature of science. As I have said, the function of a trait is the reason why it is a stable feature of things of a certain sort. That the trait is valuable can surely form no part of this explanation.

MOHAN MATTHEN

University of Alberta

---

[1] See Richard Levins and Richard Lewontin, *The Dialectical Biologist* (Cambridge: Harvard, 1985), chs. 1–3, for a forceful discussion of this point.

THE DYNAMICS OF POWER AND KNOWLEDGE IN SCIENCE*

How should we understand knowledge and power in philosophical reflections upon scientific work? Most recent philosophical and sociological interpretations of science regard knowledge and power as analytically distinct domains of human concern, which nevertheless interact in important ways. One domain includes representations of the world, and the evidence, reasons, and values that support or challenge their claims to truth or warranted assertibility. The other consists of human actions, and the cultural and institutional setting in which they affect one another causally. Their basic possible interactions are familiar: power may suppress knowledge, or distort it ideologically; it may also provide resources needed to achieve or accelerate knowledge acquisition, which in turn may augment power.

Most discussions take both knowledge and power as things acquired, possessed, and used by agents. The agent/knower is then the point of contact between knowledge and power. Power only censors or distorts knowledge by acting on knowers; knowledge only enhances or challenges power by empowering or disempowering particular agents. Such accounts locate both power and knowledge in the temporal present. Knowledge requires operant or abiding beliefs held for reasons presently available to and influential upon knowers. Power likewise operates only where actions are presently committed or omitted because of actions performed by others (including *present* threats or promises of future acts).

This paper proposes an alternative understanding of knowledge, power, and science. This alternative is designed to account for important aspects of power relations, and of the production and assessment of knowledge, which standard conceptualizations omit or misrepresent. Knowledge and power here represent not distinct domains of things, but rather interconnected ways of posing questions to and concerns about the same domain. Furthermore, this domain of knowledge/power is not limited to things present to and possessed by agent/knowers, but is temporally, socially, and spatially dispersed. Its components are irreducibly heterogeneous (knowledge includes more than beliefs or propositions; power goes beyond actions causally influencing the actions of others).

---

* To be presented in an APA symposium on Knowledge, Power, and Science, December 30. Helen E. Longino will be co-symposiast, and Philip Kitcher will comment. See this JOURNAL, this issue, 666–74 and 675–6, respectively, for their contributions.

0022-362X/91/8811/658–65

An important first step toward such a reconception occurs in Thomas Wartenburg's[1] recent account of power as dynamic and socially situated. Wartenburg argues that power is always mediated by a "social alignment": a relationship between two agents is a *power* relationship only because and to the extent that other agents will normally respond to the two agents in ways coherently aligned with the dominant agent's actions. Wartenburg's claim has several important consequences. First, power is dispersed and heterogeneous. It is not simply possessed and exercised by dominant agents, nor even located specifically within their relationships to subordinate agents. Power is mediated by the actions of many peripheral agents, which establish or enforce the connection between dominant agents' actions and the fulfillment or frustration of subordinate agents' desires. Social alignments extend far beyond the immediate power relationship, for they include not only the direct orientation of others' actions by what a dominant agent does, but also the availability of alternative access to the goods which the exercise of power would enable or deny. Large social structures are thus implicated in apparently local and specific exercises of power.

A second important consequence is that power is dynamic and temporally extended, and exists only through being reproduced:

> The *present* actions of a dominant agent count on the *future* actions of the aligned agents being similar to their *past* actions. But this faith in a future whose path can be charted entails that the dominant agent not act in a way that challenges the allegiance of his aligned agents, for only through their actions can that future be made actual (*ibid.*, p. 170).

Social alignments thus both constrain and constitute power; they also offer opportunities to resist or evade power by disconnecting its supporting alignment or forming counteralignments.

The social world thus becomes an array of overlapping social alignments oriented by ongoing struggles of domination and resistance, within which agents and their actions are situated. Wartenburg explicitly restricts the domain of power to this social world (*ibid.*, p. 3ff). But the alignments that mediate power and its resistance or evasion also include tools, processes, practices, and things worked with and upon. Power relations require not only keeping other human agents in line, but also a reliable alignment of the physical environment. Blood tests for HIV antibodies, temperature-sensitive rocket booster O-rings, photoelectric conversion of solar energy, and antimissile radars are but prominent examples of the mediations of power and resistance by physical things and processes.

---

[1] *The Forms of Power: From Domination to Transformation* (Philadelphia: Temple, 1990).

Excluding the behavior of things from alignments constituting "social power" is understandable. Marx's criticism of commodity fetishism alerted social theorists to the dangers of reifying social relations. But restricting alignments of power to relations among agents overlooks how things *resist* those alignments (they are recalcitrant, confusing, unavailable when needed, or in the way). We avoid fetishism not by strictly separating the natural and the social, nor by reducing the natural to the social, but by rethinking the distinction. *Knowledge and Power*[2] initiated such rethinking by extending "power" beyond narrowly "social" interactions to include transformations of the configuration of practices and things within which actions are intelligible. Materials, things, processes, and practices shaped by the construction and extension of scientific "microworlds" play an especially important role in such configurations, and hence in the alignment of power. Understanding power in this way neither reifies social relations, nor ascribes interests, values, rights, or a "natural" order to things. It does, however, show that distinctions between the "natural world" and the "social world" cannot legitimately confine power to the latter.[3]

Scientific knowledge can be usefully modeled upon this reinterpretation of power as dynamic and heterogeneous. Knowledge is then not something possessed and exchanged, but a process mediated by heterogeneous and contested alignments. To understand this approach, consider first the complex practical achievements that make scientific domains accessible. *Knowledge and Power* argued that things like electrical currents, retroviruses, tectonic plates, or pulsars only become possible objects of knowledge or discourse within such a complex practical field, shaped by the availability of equipment, and subtle technical and theoretical skills (*op. cit.*, chs. 4–5).

For example, "genes" emerge as objects of possible discourse through accumulations of capabilities and insights in specific contexts (e.g., laboratories with their own projects, protocols, and materials, but also experimental systems such as pisum, drosophila, maize, and bacteriophage). These cannot be extended to other locations, or related objects, without complex and subtle mutual adaptations. Laboratory practices and equipment, and the knowledge they embody, must be standardized, simplified, and adapted to new purposes, while the working environment (both material and conceptual) to which they are extended must also be modified to accommodate them.

---

[2] *Knowledge and Power: Toward a Political Philosophy of Science* (Ithaca: Cornell, 1987).

[3] See my *Knowledge and Power* on "configurations" of practices, pp. 62–5, 73–8, 182–5; on "microworlds," pp. 101–8; on power and the "social world," chs. 6–7.

These gradual transformations, reproductions, and extensions of localized knowledges also constitute alignments in a sense analogous to Wartenburg's social alignments. Some elements of a confusing array of interacting projects, practices, and capabilities reinforce and strengthen one another, and are taken up, extended, and reproduced in various new contexts, while others remain isolated from or in conflict with these emergent strategies, and gradually become forgotten or isolated curiosities. These strategic alignments are constitutive of knowledge. Thus, a statement, skill, or model does not count as knowledge in isolation or instantaneously. Rather, their epistemic standing depends upon relations to many other practices and capabilities, and especially upon how these relations are reproduced, transformed, and extended.

Knowledge therefore has a more complex temporality than do operant or abiding warranted beliefs. To take something as knowledge is to project it as a resource for ongoing activity, whether in further research, or in various "applications." Knowledge in this sense *circulates,* and its points of articulation, or even collection and assessment, are caught up in its circulation. Proposals of possible new knowledge, informally or in print, are programmatic, and depend upon interpretation and use in subsequent work. What is gathered together retrospectively is likewise oriented toward further advances, and shaped by that projection. Scientific work continually reorganizes what is known as a resource for further investigation.

Conflict is important to the temporality of scientific knowledge. If knowledge only exists through its reproduction and circulation, conflict focuses and directs that circulation. Knowledge is typically contested in specific respects, and in just those respects will it be developed and articulated most extensively and precisely. Resistance calls for new and more powerful techniques, more precise and careful measurement, and more refined theoretical models to eliminate or bypass possible inaccuracies or unrealistic assumptions. These new knowledges often, in turn, provide new directions or problems for research. Hence, whole clusters of new local capabilities and their extension into new contexts emerge from where knowledge is resisted. But the contrary also holds: where knowledge goes unchallenged, there is little or no articulation or development. And where previous resistance vanishes, knowledge also ceases to proliferate.

Scientific conflicts cannot be easily reduced to traditional epistemic categories. Obviously, resistance may focus upon gaps in data, inconsistencies or dubious assumptions in theoretical models, countervailing evidence, or conflicts with other work. But knowledge can also be resisted because the procedures and capabilities for its articulation and development are expensive, environmentally unsound, cruel to animals, politically sensitive, of too little or too much interest to the military, unprofitable, etc. Philosophers have often tried to distinguish considerations internal to knowledge from those impinging upon it from the "outside." Yet the locality of scientific work and its dynamics undercut such distinctions. The sense of a claim and its articulation and deployment in further research and development depend upon heterogeneous considerations. The small local decisions about research materials, equipment, procedures, funding, personnel, skill development, and the like shape the development of local knowledges that invest and underwrite the knowledge claims which typically concern philosophers. The justifications of these decisions typically interchange and balance supposedly internal and external considerations.

This heterogeneity of knowledge and its resistance highlights the inadequacy of conceiving knowledge as a distinct *object* domain. Instead, "knowledge" is one way of codifying and intervening in things that could also be addressed in overlapping and countervailing ways under other headings. Understanding knowledge dynamically in all its local proliferation and heterogeneity invokes considerations "properly" belonging to other domains were the world present in such discrete bundles.

Philosophers often appeal to "background knowledge" to account for these aspects of knowledge. But "background knowledge" is usually made homogeneous (typically as a field of beliefs) and cumulative.[4] Instead, I take the "field" within which scientific claims acquire significance and justification to include more than a web of belief: skills and techniques, instruments and material systems (including networks of manufacture and supply), resource availability (money and facilities, but also staff, information, an audience, etc.), institutional structures, relevance to other social practices or political concerns, and much more. The background against which knowledge claims are interpreted, used, and assessed also varies across the contexts in which they appear. The "same" statement (technique, device, etc.) is often employed differently in different disciplines, but even within disciplines there is considerable local variability (and conflict!) in what counts as epistemically relevant background. Such backgrounds also change over time, especially as research shifts direction. The frequent invocation of "holism" to indicate the field structure of knowledge is thus misleading; it suggests a unity, coherence, and systematicity belied by the dynamic character of scientific practice. But my emphasis on the disunity and heterogeneity of scientific knowledge is not skeptical or debunking: partial communication is a normal circumstance of scientific work, and can itself be epistemically productive.[5]

---

[4] One sophisticated attempt to shoehorn such heterogeneous considerations into a unified field of "background knowledge" is Dudley Shapere's *Reason and the Search for Knowledge* (Boston: Reidel, 1984).

If knowledge and power should each be reconceived as dynamic, heterogenous, and dispersed across complex social and material "alignments," and neither demarcates a distinct object domain, how might this affect *relations* between knowledge and power in the sciences? First, the traditional examples of knowledge/power relations (censorship, ideology, and resource provision) are no longer merely marginal issues either for epistemology or for political thought. A genuinely dynamic account of knowledge allows that censorship or ideology can become deformations of knowledge itself (not merely constitutive of pretenses to knowledge). There is no sovereign standpoint *outside* the ongoing conflicts over knowledge and truth from which to distinguish false consensus and knowledge (this claim is antithetical to relativism, which itself takes up a "sovereign" standpoint in assigning all knowledge claims equal adequacy); hence, the criticism of censorship and ideology is internal to epistemology and philosophy of science rather than just doing the boundary work of keeping knowledge assessment "pure."

Likewise, the ways in which power and knowledge provide one another with material resources are partially constitutive of each. Complex and sophisticated instrumentation, the services of skilled personnel, institutional support, and the other resources provided via access to power are not merely instrumental to knowledge. Knowledge is embodied in skills, techniques, and machines, and its development is directed as much by how objects can be manipulated and surveyed as by how they are represented. Understanding the power relations shaping possibilities for action also requires taking account of how knowledge helps constitute and reshape the alignments through which power is exercised.

---

[5] On the partiality of scientific communication, and its implications for knowledge, see Steve Fuller, *Social Epistemology* (Bloomington: Indiana UP, 1988).

But at least two important issues for understanding power, knowledge, and science are not readily visible at all in traditional terms. First, the very attempt to make a domain of objects more readily or precisely knowable may realign power. Second, social movements or transformations may have simultaneously epistemic and political consequences.

The first case was discussed extensively in *Knowledge and Power*. Scientific knowledge of new objects or situations is acquired by developing new ways of manipulating, partitioning, surveying, and tracking those things. These, in turn, require new practices of control (often *self*-control) of the behavior of scientific researchers. Extending knowledge beyond the research site typically also requires extending and adapting specific materials, practices, and disciplines. Such extensions are frequently considered "engineering" (and eventually routine procedure and maintenance) rather than science; yet epistemologically, they fulfill the crucial expectation that knowledge is not an artifact of the research context. In this way, they constitute scientific work as ***knowledge*** of the world. Yet this very extension and validation of knowledge simultaneously disciplines and controls the "action-environment" of many human agents.[6]

I have also discussed the second issue elsewhere,[7] through a comparative discussion of creationist biology and multidisciplinary feminist scholarship. Creationist strategies were an interesting mix of trying to subvert or co-opt elements of the dominant epistemic alignment which established and enforced the rationality of belief in Darwinian theory, and trying to create alternative alignments (Christian schools, creationist research and textbooks, etc.) to bypass it. Feminist scholarship has also typically built new research programs that combined new arguments and evidence, a resourceful appropriation and redirection of elements of orthodox scholarship, and the establishment of alternative "social" alignments to reduce their dependence upon unregenerate opponents (feminist presses and journals, professional associations and graduate programs, and political alliances). For both, the boundary between knowledge and politics or religion is at issue in the very claims to knowledge that emerge from these new epistemic/political alignments. And the point of my argument was to show that one cannot, for example, adequately account for the epistemic successes of feminist research, and the substantial failure of the creationist program, without taking seriously the inseparability of issues of epistemic warrant and political power.

---

[6] Wartenburg (*op. cit.*, p. 80–4) uses this term for agents' understanding of their situation as oriented toward possible courses of action.

[7] "Foucault and the Natural Sciences," forthcoming

Even those inclined to accept my arguments might still suspect that the intertwining of knowledge and power is only marginal to science. But, besides the growing literature tracing gender throughout scientific work, this suggestion is also effectively challenged by Simon Schaffer's and Steven Shapin's[8] study of the seventeenth-century development of norms of experimental practice, and by Sharon Traweek's[9] forthcoming study of the scientific and political marginality of particle physics in Japan. Schaffer and Shapin depict experimental methodology as itself the product of a sociopolitical movement (institutionally embodied in the Royal Society) in the same way we might describe creationist biology or feminist anthropology as produced by sociopolitical movements. Traweek's rich and complex study also shows how this intertwining of epistemic and sociopolitical alignments is not merely a historical moment long surpassed by the development of science.

But Shapin and Schaffer importantly misdescribe their own achievement. Their conclusion that "Hobbes was right" (*op. cit.* p. 344) mistakenly suggests reducing epistemic considerations to sociopolitical ones.[10] Whereas their argument shows only that epistemic and political questions do not belong to distinct domains of inquiry, and that answering each requires sustained attention to the other. Knowledge and power are neither two different kinds of thing, nor the same thing (no matter which direction the reduction allegedly goes). They represent different ways of configuring and interrogating a wide range of our engagements with the world and each other. If each field of questions has its own integrity and significance, reflection on the dynamics of knowledge and power suggest nevertheless that neither is adequately addressed without serious attention to the other.

JOSEPH ROUSE

Wesleyan University

---

[8] *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life* (Princeton: University Press, 1985).

[9] "Border Crossings," in Andrew Pickering, ed., *Science as Practice and Culture* (Chicago: University Press, forthcoming).

[10] Shapin and Schaffer have been similarly criticized by Bruno Latour, "Postmodern? No, Simply Amodern! Steps Towards an Anthropology of Science," *Studies in the History and Philosophy of Science*, XXI (1990): 145–71.

MULTIPLYING SUBJECTS AND THE DIFFUSION OF POWER*

Knowledge *is* power. Pundits from Bacon to Michel Foucault have told us this about scientific knowledge. Bacon understood that knowledge of causes (as long as these were accessible) enabled the knower to intervene in and control natural processes. Foucault, speaking to generations much more cynical about the results of such knowledge, urged us to see how power/knowledge is productive: phenomena are created (come into view as phenomena) by our knowledge of them. And surely we feel the identification of knowledge and power in "smart bombs," "thinking machines," and "designer genes."

The sciences have transformed conditions of life in industrialized portions of the world, both conceptually as models of knowledge and materially through science-based technologies. Why, then, do some of us feel so uneasy about the identification of knowledge and power? There are surely various sources for and locations of this uneasiness. Those of us who are feminists have been struck by the interlocking character of several aspects of knowledge and power in the sciences. Women have been excluded from the practice of science, even as scientific inquiry gets described both as a masculine activity and as demonstrating women's unsuitability to engage in it, whether because of our allegedly deficient mathematical abilities or our insufficient independence. Some of us notice the location of women in the production of the artifacts made possible by new knowledge: swift and nimble fingers on the microelectronics assembly line. Others notice the neglect of women's distinctive health issues by the biomedical sciences, even as new techniques for preserving the fetuses they carry are introduced into hospital delivery rooms. The sciences become even more suspect as analysis of their metaphors (for example, in cell biology and in microbiology) reveals an acceptance (and hence reinforcement) of the cultural identification of the male with activity and of the female with passivity. Finally, the persistent privileging of explanatory models constructed around relations of unidirectional control over models constructed around relations of interdependence, a privileging that mirrors social relations, has prompted feminist critics to question the very idea

---

* To be presented in an APA symposium on Knowledge, Power, and Science, December 30. Joseph Rouse will be co-symposiast, and Philip Kitcher will comment. See this JOURNAL, this issue, 658–65 and 675–6, respectively, for their contributions.

0022-362X/91/8811/666–74

of a scientific method capable of adjudicating the truth or probability of theories in a value-neutral way.

While the sciences have increased human power over natural processes, they have done so in a lopsided way, according to this analysis, systematically perpetuating women's cognitive and political disempowerment (as well as that of other groups marginalized in relation to the Euro-American drama). One obvious question, then, is whether this appropriation of power is an intrinsic feature of science or whether it is an incidental feature of the sciences as practiced in the modern period, a feature deriving from the social structures within which the sciences have developed. Feminists have answered this question in a number of ways. I shall review some of these before outlining my own answer.

I FEMINIST EPISTEMOLOGICAL STRATEGIES

1. CHANGING THE SUBJECT

Most traditional philosophy of science (with the problematic exception of Descartes's) has adopted some form of empiricism. Empiricism's silent partner has been a theory of the subject, that is, of the knower. The paradigm knower in Western epistemology is an individual—an individual who, in several classic instances, has struggled to free himself from the distortions in understanding and perception which result from attachment. For Plato, for example, knowledge of the good is possible only for those whose reason is capable of controlling their appetites and passions, some of which have their source in bodily needs and pleasures, others of which have their source in our relations with others. The struggle for epistemic autonomy is even starker for Descartes, who suspends belief in all but his own existence, in order to recreate a body of knowledge cleansed of faults, impurities, and uncertainties. For Descartes, only those grounds available to a single, unattached, disembodied mind are acceptable principles for the construction of a system of beliefs. Most subsequent epistemology has granted Descartes's conditions and disputed what those grounds are and whether any proposed grounds are sufficient grounds for knowledge. Descartes's creation of the radically and in principle isolated individual as the ideal epistemic agent has for the most part gone unnoticed.[1] The individual consciousness that is the subject of knowledge is transparent to itself, operates according to principles that are independent of embodied experience, and generates knowledge in a value-neutral way.

---

[1] The later philosophy of Wittgenstein does challenge the individualist ideal. Until recently few commentators have developed the anti-individualist implications of his work. See Naomi Scheman, "Individualism and the Objects of Psychology," in Sandra Harding and Merrill Hintikka, eds., *Discovering Reality* (Boston: Reidel, 1983), pp. 225–44

Feminist epistemological strategies, while sometimes described as a rejection of empiricism, can also, and perhaps better, be described as changing the subject. Feminist empiricism (as described by Sandra Harding[2]) identifies offending science as the product of male bias, and proposes replacing the male (or androcentric) subject of knowledge with an unbiased subject. On this view, certain areas of science, having to do with sex and gender, are deformed by gender ideology, but the methods of science are not themselves masculinist and can be used to correct the errors produced by ideology. The ideal knower is still the purified mind, and epistemic or cognitive authority inheres in this purity. This strategy, as Harding has observed, is not effective against those research programs which feminists find troublesome but which cannot be faulted by reference to the standard methodological precepts of scientific inquiry.

Another approach is, therefore, the standpoint approach. There is no one position from which value-free knowledge can be developed, but some positions are better than others. Standpoint epistemologies identify a particular social location as epistemically privileged.[3] Traditional Marxist epistemologies identify the standpoint of the proletariat; feminists identify the standpoint of women. By valorizing the perspectives uniquely available to those who are socially disadvantaged, standpoint theorists turn the table on traditional epistemology: the powerless are those with epistemic legitimacy, even if they lack the power that could turn that legitimacy into authority. The difficulties of the standpoint approach come into high relief when it is a women's or a feminist standpoint that is in question. Women occupy many social locations in a racially and economically stratified society. A social theory is therefore needed to ascertain which of these locations is the epistemologically privileged one. But in a standpoint epistemology, a standpoint is needed to justify such a theory. Most standpoint theorists locate the epistemic advantage in the social experience of the oppressed whose perspective they champion. A similar change of subject is proposed by those identifying the problems with science as a function of the psychodynamics of individuation. Evelyn Fox Keller,[4] for example, has understood the traditional ideal of scientific objectivity as the ideal of detachment of the scientist from the object of study. Epistemic and affective ideals are intermingled and, from the psychoanalytic perspective she adopts, distorted affective development—autonomy as exaggerated separateness—has produced a distorted epistemic ideal—objectivity as radical detachment. She has, therefore, proposed an alternative conceptualization of autonomy. She has contrasted static autonomy with what she calls dynamic autonomy, an ability to move in and out of intimate connection with the world. Dynamic autonomy provides the emotional substructure for an alternative conception of objectivity: dynamic objectivity. Keller analyzes static objectivity as generating and satisfied by accounts of natural processes that foreground controlling relationships. The knower characterized by dynamic objectivity, by contrast, does not seek power over phenomena but acknowledges instead the ways in which knower and phenomena are in relationship, as well as the ways in which phenomena themselves are complexly interdependent. While either changing or diffusing power, all these strategies have in common a focus on the individual epistemic agent, on the autonomous subject.

---

[2] *The Science Question in Feminism* (Ithaca: Cornell, 1986).

[3] Cf. Nancy Hartsock, "The Feminist Standpoint: Developing the Ground for a Specifically Feminist Historical Materialism," in *Discovering Reality*, pp. 283–310.

II FEMINIST EPISTEMOLOGICAL STRATEGIES 2. MULTIPLYING SUBJECTS

Developments in the philosophy of science suggest the individual knower is an inappropriate focus for the purpose of understanding (and changing) science. The natural sciences are characterized in the traditional view by a methodology that purifies scientific knowledge of distortions produced by scientists' social and personal allegiances. The essential features of this methodology—explored in great detail by positivist philosophers of science—are observation and logic. Much philosophy of science in the last twenty-five years has been preoccupied with two potential challenges to this picture of scientific methodology—the claim of Thomas Kuhn, Paul Feyerabend, and N. R. Hanson that observation is theory-laden, and the claim of Pierre Duhem that theories are underdetermined by data. One claim challenges the stability of observations themselves, the other the stability of evidential relations. Both accounts have seemed to permit the unrestrained expression of scientists' subjective preferences in the content of science. If observation is theory-laden, then observation cannot serve as an independent constraint on theories, thus permitting subjective elements to constrain theory choice. Similarly, if observations acquire evidential relevance only in the context of a set of assumptions, a relevance that changes with a suitable change in assumptions, then it is not clear what protects theory choice from subjective elements hidden in background assumptions. As long as the scientific knower is conceived of as an individual, knowing best when freed from external influences and attachment, that is, when detached or free from her context, the puzzles introduced by the theory ladenness of observation and the dependence of evidential relations on background assumptions will remain unsolved.

---

[4] *Reflections on Gender and Science* (New Haven: Yale, 1984).

Applying the feminist insight that we are all in relations of interdependence, I have suggested that scientific knowledge is constructed not by individuals applying a method, but by individuals in interaction with one another in ways that modify their observations, theories and hypotheses, and patterns of reasoning. Scientific method, thus, includes not only the complex of activities that constitute hypothesis testing through comparison with experiential data, in principle an activity of individuals. It involves equally centrally the subjection of hypotheses and the background assumptions in light of which they seem to be supported by data to varieties of conceptual and evidential criticism.[5] Because background assumptions can be and most frequently are invisible to the members of the scientific community for which they are background, and because unreflective acceptance of such assumptions can come to define what it is to be a member of such a community (thus making criticism impossible), effective criticism of background assumptions requires the presence and expression of alternative points of view. This sort of account allows us to see how social values and interests can become enshrined in otherwise acceptable research programs, i.e., research programs that strive for empirical adequacy and engage in criticism. As long as representatives of alternative points of view are not included in the community, shared values will not be identified as shaping observation or reasoning.

Scientific knowledge, on this view, is an outcome of the critical dialogue in which individuals and groups holding different points of view engage with each other. It is constructed not by individuals but by an interactive dialogic community. A community's practice of inquiry is productive of knowledge to the extent it facilitates transformative criticism. The constitution of the scientific community is crucial to this end as are the interrelations among its members. Community level criteria can, therefore, be invoked to discriminate among the products of scientific communities, even though context-independent standards of justification are not attainable. At least four criteria can be identified as necessary to achieve the transformative dimension of critical discourse:

---

[5] For argument for and exposition of these points, see my *Science as Social Knowledge* (Princeton: University Press, 1990).

(1) There must be publicly recognized forums for the criticism of evidence, of methods, and of assumptions and reasoning.
(2) The community must not merely tolerate dissent, but its beliefs and theories must change over time in response to the critical discourse taking place within it.
(3) There must be publicly recognized standards by appeal to which criticism is made relevant to the goals of the inquiring community. (While there need not and probably is not a set of standards common to all communities, the general family of standards includes such cognitive virtues as accuracy, coherence, and breadth of scope, and social virtues such as fulfilling technical needs.)
(4) Finally, objective communities must be characterized by equality of intellectual authority. What consensus exists must be the result not of the exercise of political or economic power, but a result of critical dialogue in which all relevant perspectives are represented.

Rather than a view from nowhere or from somewhere, this notion of objectivity idealizes the view from everywhere. While there are any number of objections advocates of such a notion must address, I shall confine myself here to one major problem.

III. DILEMMAS OF PLURALISM

This sort of account is subject to the following dilemma.[6] What gets produced as knowledge depends on the consensus reached in the scientific community. For it to count as objective, the community must be adequately diverse. How is scientific knowledge possible while pursuing socially constituted objectivity? That is, (1) if objectivity requires pluralism in the community, then scientific knowledge becomes elusive; but (2) if consensus is pursued, it will be at the cost of quieting critical oppositional positions.

My strategy for avoiding this dilemma is to detach scientific knowledge from consensus, if consensus means agreement of the entire scientific community regarding the truth or acceptability of a given theory. I suggest that we look at the aims of inquiry (at least some) as satisfied by embracing multiple and, in cases, incompatible theories that satisfy local standards. This detachment of knowledge from universal consensus can be made more palatable than it might first appear by two moves. One of these is implicit in treating science as a practice or set of practices; the other involves taking up some version of a semantic or model-theoretic theory of theories.

---

[6] Thanks to Sandra Mitchell for this formulation.

Beginning with the second of these, let me sketch the relevant aspects and implications of the semantic view.[7] If we take such a view, we understand a theory as a specification of a set of relations among objects or processes characterized in a fairly abstract way. Another characterization would be that, on the semantic view, a theory is the specification of a structure. The adequacy of a theory so conceived is determined by our being able to map some subset of these relations/structures on to some portion of the experienced world. Now the portions of the world stand in many relations to many other portions. Any given model or schema will necessarily select among those relations. So its adequacy is not just a function of isomorphism of one of the interpretations of the theory with a portion of the world, but that the relations it picks out are ones in which we are interested. A model guides our interactions with and interventions in the world. We want models that guide the interactions and interventions we seek. Given that different subcommunities within the larger scientific community may be interested in different relations, or that they may be interested in objects under different descriptions, different models (that if taken as claims about an underlying reality would be incompatible) may well be equally adequate and provide knowledge, in the sense of an ability to direct our interactions and interventions, even in the absence of a general consensus as to what is important. Knowledge is not detached from knowers in a set of propositions but consists in our ability to understand the structural features of a model and to apply it to some particular portion of the world. It is knowledge of that portion of the world through its structuring by the model we use.

The second move to escape the dilemma develops some consequences of treating science as practice. There are two worth mentioning. If we understand science as practice, we understand inquiry as ongoing, that is, we give up the idea that there is a terminus of inquiry which just is the set of truths about the world. (What La-

---

[7] My understanding of the semantic view is shaped by its presentations in Bas van Fraassen, *The Scientific Image* (New York: Oxford, 1980); and in Ronald Giere, *Explaining Science* (Chicago: University Press, 1988), as well as by conversations with Richard Grandy and Elisabeth Lloyd

Place's demon knew, for example.) Scientific knowledge from this perspective is not the static end point of inquiry, but a cognitive or intellectual expression of an ongoing interaction with our natural and social environments. Scientific knowledge, then, is a body of theories and their articulations onto the world that change over time in response to the changing cognitive needs of those who develop and use the theories, in response to the new questions and anomalous empirical data revealed by applying theories, and in response to changes in associated theories. Both linear-reductionist and interactionist models reveal aspects of natural processes, some common to both, some uniquely describable within the terms proper to one but not both sorts of model. If we recognize the partiality of theories, as we can when we treat them as models, we can recognize pluralism in the community as one of the conditions for the continued development of scientific knowledge in this sense.

Secondly, rejecting the notion of a final terminus to inquiry allows us to appreciate another feature of treating science as practice. Practices have aims, and as such, are evaluated in terms of their ability to realize the goals of the practitioners who engage in them. But there is a fundamental tension between two central aims of the practice of inquiry: the explanatory aim and the truth-seeking aim. In pursuing explanation, we seek theories of generality and scope and mutual coherence by extending features of a model from one domain to another. In pursuing truth, we subject the constituent statements of a theory to empirical test. Rather than extensions of the model we seek to expose faults in the model. The explanatory aim is satisfied by consensus-formation practices while the truth-seeking aim is satisfied by critical practices. Thus, if the tension is between mutually exclusive aims, the dilemma constructed on the horns of consensus and pluralism is a constitutive tension of scientific inquiry, a tension which, if the remarks in the preceding paragraph are correct, can be understood as one of the engines of scientific change.

IV. CONCLUSIONS

Understanding scientific knowledge in this way supports at least two further reflections on knowledge and power. First of all, the tension between explanatory and truth-seeking aims suggests that issues concerning the constitution of the scientific community will be ongoing practical and political problems. The need for models within which we can situate ourselves and the interactions we desire with the natural world will militate against the inclusiveness required for an adequate critical practice, if only because the elaboration of any model requires a substantial commitment of material and intellectual resources on the part of a community.[8] This means that, in a power stratified society, the inclusion of the less powerful, and hence of models that could serve as a resource for criticism of the received wisdom in the community of science, will always be a matter of conflict. At the same time, the demand for inclusiveness should not be taken to mean that every alternative view is equally deserving of attention. There is always a danger that the politically marginal will be conflated with the crackpot. While one function of public and common standards is to remind us of that distinction, and to help us draw it in particular cases, I do not know of any simple or formulaic solution to this problem.

Those critiques of scientific epistemology which urge a change of subject preserve the structures of cognitive authority but propose replacing those currently wielding authority with others: a genuinely unbiased subject in one case, a differently located or a differently formed subject in the other. Either no assumptions or different assumptions will be engaged in the knowledge constructive process. In the view I am advocating, which makes salient features of knowledge construction made invisible by more traditional accounts, the structures of cognitive authority themselves must change. No segment of the community, whether powerful or powerless, can claim epistemic privilege. If we can see our way to the dissolution of those structures, then we need not understand the appropriation of power in the form of cognitive authority as intrinsic to science. Nevertheless, the creation of cognitive democracy, of democratic science, is as much a matter of conflict and hope as is the creation of political democracy.

HELEN E. LONGINO

Rice University

---

[8] For a somewhat different approach to a similar question, see Philip Kitcher, "The Division of Cognitive Labor," this JOURNAL, LXXXVII, 1 (January 1990): 5–23.

SOCIALIZING KNOWLEDGE*

We are accustomed to thinking of propositional knowledge as some kind of true belief: we can be ecumenical by tolerating vagueness and suppose that knowledge is "well-grounded" true belief. So conceived, propositional knowledge is the property of individuals, those who believe the true things in the appropriate well-grounded way. Yet, as historians and sociologists of science have made abundantly clear, scientific knowledge is a social achievement. Individual knowers are epistemically dependent on those around them.

There is a relatively conservative way of reacting to this news. Apologizing for their focus on individual reasoning rather than social interaction, epistemologists can acknowledge the neglected problem of looking at the ways in which social exchanges, reliance on others, and particular structures of authority and deference relate to the well-grounding of belief. The traditional account of knowledge is retained, and knowledge is socialized by recognizing the need to understand those social conditions which promote the well-groundedness of individual belief.[1]

Historians and sociologists typically want more concessions, and the papers of Joseph Rouse and Helen E. Longino present a far more radical approach to socializing epistemology, one that is in tune with, if not inspired by, recent work in the history and sociology of science.[2] Their writings suggest a family of conceptions of knowledge that break with the traditional approach. Neither Longino nor Rouse offers a fully explicit rival analysis of knowledge. I believe that exploring the possibilities is essential for evaluating their claims.

Despite the talk of knowledge "circulating," neither author offers any coherent alternative to the notion that knowledge involves individual belief. To be sure, an individual's ability to entertain a particular proposition may depend on her relations to others—but that is a

---

* Abstract of a paper to be read in an APA symposium on Knowledge, Power, and Science, December 30, commenting on papers by Joseph Rouse and Helen E. Longino. See this JOURNAL, this issue, 658–65 and 666–74, respectively, for their contributions.

[1] For examples of this relatively conservative response, see Alvin Goldman, "Foundations of Social Epistemics," *Synthese*, LXXIII (1987):109–44; John Hardwig, "Epistemic Dependence," this JOURNAL, LXXXII, 7 (July 1985): 335–49; my "The Division of Cognitive Labor," this JOURNAL, LXXXVII, 1 (January 1990): 5–22, and "Authority, Deference, and the Role of Individual Reasoning in Science," in Ernan McMullin, ed., *The Social Dimension of Scientific Knowledge* (Notre Dame: University Press, 1991).

[2] The connections are very clear in the important books that stand behind their papers: Rouse's *Knowledge and Power* (Ithaca: Cornell, 1987); and Longino's *Science as Social Knowledge* (Princeton: University Press, 1990).

0022-362X/91/8811/675–6

relatively familiar point.[3] The novelties in Rouse's and Longino's approaches to knowledge lie in their suggestions of surrogates for truth and well-groundedness.

Here are some alternatives to the condition that *p* (the proposition known) must be true:

(a) All members of the subject's society believe that *p*.
(b) Belief that *p* is effective in advancing the subject's attainment of a goal validated by members of the subject's society.

Here is an alternative to the condition that the belief that *p* must be well-grounded:

(c) The subject's belief is produced in a way that meets the epistemic standards endorsed in the subject's society.

Finally, any of (a)–(c) may be subjected to the condition:

(d) The subject's society must be epistemically well-ordered (for example, it must honor and actualize the ideal of critical debate).

Simply substituting (a) for the traditional conditions of truth and well-groundedness identifies knowledge with consensual belief, and invites the rejoinder that that is no notion of *knowledge* at all. Both Rouse and Longino are much more subtle, favoring conditions akin to (b), (c), and (d). It is easy to see how these conditions enable Rouse to draw connections between knowledge and power and Longino to suggest that knowledge is intrinsically value-laden.

But why should we break with tradition at all? Why not socialize knowledge in the relatively conservative way I indicated above? I trace the motivation for radical departures to the thesis that truth and well-groundedness are inaccessible. I shall attempt to argue that this thesis is mistaken and that, if taken seriously, it spells trouble for the surrogate notions as well.

Finally, if we adopt the surrogate notions, can we reclaim any type of objectivity? If our notion of objectivity includes—as I believe it should—the idea that there are rival, epistemically well-ordered, societies and socially approved beliefs within each of them that differ, and that some of these beliefs are epistemically superior to others, then objectivity cannot be recaptured. (So, at least, I shall try to show). Rouse and Longino offer conceptions of knowledge that are far more sophisticated than village relativism but, in the end, they seem to be forced to embrace all the consequences that make relativism so unappealing.

PHILIP KITCHER

University of California/San Diego

---

[3] The *locus classicus* is Hilary Putnam's idea of "division of linguistic labor" in his "Meaning and Reference," this JOURNAL, LXX, 19 (November 8, 1973): 699-711.

WHY BOTHER:
IS LIFE WORTH LIVING?*

We must beware of our penchant to dismiss the cliche phrase, especially posed as a seemingly trite rhetorical question. At first glance, the query as to whether life is worth living strikes us as somewhat routinely jocular, a sort of throw-away question to which one would not expect a reply, let alone an answer. Nonetheless, if we take the question at dead reckoning (pun intended, for the inquiry, after all, is about death, that is, my death) then its seriousness leaps to the fore.

THE QUESTION

So, let us ask ourselves, each in turn, is *my* life worth living? Surely, we must avoid a present-minded response for at some time, the *yin* and *yang* in most of our lives, the only appropriate refrain would have been to wit: I wish I could die.

Yet, even if we extricate ourselves from an immediate personal crisis, an identical plea, justifiably, could be entered. The arrangement of my own death, despite its chilling finality, does bring with it two salient advantages. First, it is a free act and, for that matter, perhaps my only free act, ever. Second, its very finality is also liberating as in the sadly bravado comment, "Well, I shall not have to put up with that (or anything) anymore."

If the reader-listener finds this too lachrymose or riddled with the allegedly nasty vice of systemic despair, I counter by offering a brief diagnosis of what it is to "do" living.

THE SETTING

The first, foremost, and permanent ontological fact of our human situation is that we were born to live but sure to die. The awareness of this central, irreducible, incontrovertible fact comes to each of us at different times in our lives. For me, being full-blood Irish-American, the awareness came very early, at age four or five. The setting was traditional, no less than participation in that glorious linguistic piece of self-deception, the "wake." And just why, I would boldly ask, is Aunt Peggy asleep, here in the parlor? I was told that this sleep was necessary if she were to "wake" again, as in the powerful and oft-healing biblical motif, paraphrased as "if the seed does not die, ye shall not have eternal life."

---

* To be presented in an APA symposium on What Makes Life Worth Living, December 30. Karsten Harries will be co-symposiast, and Jay Rosenberg will comment. See this JOURNAL, this issue, 684–90, for Harries's contribution.

0022-362X/91/8811/677–83

I came to accept that explanation, albeit with the canny dubiety often found in the not so virginal mind of a child. And, the click of tall glasses filled with rye and ginger created a domestic celebratory atmosphere that could only bode well for the future of Aunt Peggy. It was much later in my life that I became sadly aware that the alcohol was far less a celebratory liquid amulet than an attempt to deaden intractable pain. The searing truth of this reached a crescendo when I drank my way through my father's wake, a pain still throbbing.

Some among us believe that in time we shall be redeemed. I do not so believe. Some among us believe that death is *au naturel,* or as Marcus Aurelius contends, a sort of spiritual nitrogen cycle, knit by an inexorable and holy bond. I do not so believe. Still others refer to us in our future state of finality as "the grateful dead." I do not so believe.

Neither you nor I were asked if we were willing to start on this journey. Our conception resulted from the acts of others. Is it not plausible, then, that we may come to resent being told that we must finish a journey begun tychastically by us, willy-nilly as it were. Our opponents charge us with *hubris,* an important Greek word which means pride, nay stubborn, self-destructive pride as testated by Achilles sulking in his tent.

My version of *hubris* is that it means the acting out of a perpetual adolescent protest. Yet, is *hubris* an accurate description of those of us who have serious doubt about whether life is worth living and who refuse to accept a supine gentility as the appropriate response to mortality? I think not. To the contrary, and paradoxically, I offer that only a response of refusal to accept the righteous character of the inevitability of death can make it possible for life to be worth living.

THE PROBLEMATIC

So much for a general philosophical landscaping. Let us now come closer to hand by means of a phenomenological diagnosis of our "text," namely, indeed, is life worth living? The term 'life' can be only used retrospectively as in my life, or, if you knew of my life, or, what a life I have had, or, I have no life, or, plaintively, this is a life? The term 'living', on the other hand, is a process word, a present participle, a happening, and a witness to the "specious present." Following John Dewey, we live only at this time and at no other time and so life is an abstraction, perched above our living as a desperate effort to identify ourselves, to become existentially instantiated. Yet, the only part of our past which exists is the past that is present to us. And the future is but a "gossamer wing," vapid, elusive, and most always, by far, doomed to be different from the futured intentions of our present.

Our living is not out of something or in something. Nor is it about something, or on behalf of something. Rather, our living is constitutive of our person. Who we are at any moment is precisely our living. Obviously, spatial metaphors fail us here. Analogous to our physiological activity, which goes on always until it stops, forever, so too does our personal living proceed inexorably, without respite, without any extrapolation for purpose of an objective view; for that, too, is our living. Our language, especially in its fidelity to a subtle yet pervasive Aristotelean bequest, leads us to believe that stasis, substance, place, thing are where we are, what we hold as if we moved from box to box, external to the flow that courses through us. My version emulates the female body, that is, I see us as uterine, a permeable membrane, eating, breathing, livering all the while. For me living is a journey, the origin of which is not of our making. The goals are *en passant* and the end is ontologically tragic, although for many of us, alas, it may be salutary, even unfortunately welcome.

THE JOURNEY

If it is so that the meaning of the last, ultimate end is unknowable, then at best our human journey involves risk. More, the ultimate end, should there be such, seemingly has no significance for me, personally, at all. The philosophical question as to the "beginning" and the "end," if there be either, and the still deeper question, "Why is there something rather than nothing?" are precisely that, questions; fascinating, intriguing, nagging, perplexing in the sense of Maimonides, yet irresolute. Their obduracy is not due to our failure, thus far, to unlock the argument, but rather to being cast in such a way that however we turn we face an experiential surd. Whatever else one wishes to say about Parmenides, he is right when he cautions us about the attempt to discuss *nothing*. In roughhouse terms, with regard to the ultimate meaning of existence we do not and can not have a clue. Does it not follow that a homely question emerges, namely, why bother?

Yes, why bother? Why *should* we live it through to our end? This is a revealingly different question than why *do* we live it through. The answers to the latter question are as old as human consciousness. They appear as brilliant constructs, anthropologically, liturgically articulated, and often accompanied by repressive totems, warnings, advice and claims, none of which has any ultimate certitude.

If I do not know the ultimate meaning of when, of how, of why, and I am deeply skeptical of the integrity of those efforts to explicate those questions so that I shall behave as cajoled, demanded or in response to a promise, the most fake of all the attempted resolutions, then again our question comes to the fore, why bother?

It may justly be asked that, if one does not bother, what else does one do? The answer is very singular and clear. One should commit suicide. Granted that this decision has to be a careful one for it is the final insulating cut, the permanent withdrawal from the process, from the fray and the end of hope for a way back in and out. We have no way of knowing whether those among us who have committed suicide have actually made a liberating decision, for the famous and infamous suicide note is always anticipatory. There can be no reportage on the aftermath from the doer of the deed. The assumption here is that the person committing suicide is fully aware of what he is doing. It is not true that suicides are "crazy" people, for the latter rarely kill themselves. It is also not true that suicide is necessarily a selfish act in that it leaves behind gaping wounds in the lives of others. Obviously, that can be and often is the scenario. Just as often, however, suicide is an act of moral courage and altruism, putting an end to the mayhem and hurt caused by the person who no longer believes that life is worth living. *Straight out then, living should be a personal choice made over against the existential, viable, often plausible, and certainly liberating option of suicide.* Certainly, to choose to go on living should be more than a response to those quietest phrases: 'life goes on' and 'so it goes', 'round and round we have it', as 'the world turns'.

As of this moment, I am living. At one point in the recent past, I chose not to go on living. Before my decision was consummated, I was personally seized and forced to reconsider. Alike with Dax Cowart, the Galveston burn victim, I have mixed feelings about that reprieve. Still, having it granted, I now undertake to ground the decision to go on living, hoping to answer the question, "Is life worth living?" affirmatively.

The sentence for which I reach to assist me does not come from the *Bible* or from the *Tao te Ching* or any other spiritual literature. It is from *The Myth of Sisyphus* by Albert Camus and reads, "I want to know if I can live with what I know and only with that." An understanding of the life and thought of Camus, especially as found in his early North African essays, makes it obvious that, for him, "to know" is not to be construed in a narrow, traditionally epistemological way. Knowing for Camus is ringed with ambience and is inclusive of the tacit, the inchoate. In my version, personal knowing is best found in our affective experiences, which no matter how dangerous, how trivial, never lie. As we feel so do we know. Concepts, set in the brilliant schemata of the philosophers, trail our percepts, wooing them into coherence, correspondence or with Ockham, *flatus voci*. As with William James, when faced with the problem of living I lean to "knowledge by acquaintance" rather than "knowledge about," unless one means to be up and about or hanging about.

THE JOURNEY: SUFFERING AS TEXTURE

Now as I decide to bother, both by acquaintance and by about, I ask myself the question, just what is it that is coming at me as I am living in the world. Is there a message, an *utter*ance, a voice from the *logoi* of persons, nature, things, and artifacts? As I listen, I hear, for the most part, the voice of unrequited suffering. If one does not believe, as I do not, that in the long run, *sub specie aeternitatis,* that all will be well, be one, be redeemed, or in the die-cast metaphysics of Las Vegas, even out, then the voice of the unrequited takes on a shrill, chilling, razor-sharp quality. The voice threatens to unseat any and all efforts at equanimity, let alone serenity.

Those among us who are reflectively knowledgeable are all too aware of the grisly fact that, since the beginning of human history, most persons did the living in a grinding, precarious, and repressed setting. In this century alone, yet to close, despite (or because of) its exquisite technical and scientific accomplishments, more than fifty million people have met death prematurely, violently, vulgarly, even obscenely at the hands and minds of other human beings, chasing one absolute or another due to religion, region, ethnicity, language, skin color, or race. Has life been worth living for those who for decades have suffered under periodic eruptions of attempted genocide, local internecine strife, or as featured by the twentieth-century, world-wide narcissistic bathing in rampant destruction?

These disasters have not befallen me, thus far. And so, is this jeremiad on behalf of collective, long-standing, seemingly intractable suffering, any skin off my nose? Question! To what extent should I allow the suffering of others affect my version of life's being worth living? The responses to this question are very revelatory as to how we understand ourselves, have ourselves as it were, as a human living creature, a person. (The ongoing destruction of our ecosystem and our treatment of all living creatures is surely germane to our question but is not of reach in the present context.)

One response, callous, but widespread, is that the suffering of others is neither our problem nor our concern. A second response, delusive, self-deceptive for me, is that there but for the grace of whatever, whomever, go I. This is a sort of talismanic approach in which the Greek notion of *moira* is transformed into the American notion of luck. Of this, I offer that Descartes was closer to the truth when he flirted with the possible existence of a *mal genie*. The Manicheans were closer still when they posited good and evil in an eternal, irresolute embrace. But did not the Calvinists, terrifying though they be, have it right when they held that some are saved and some are not and the reason is arbitrary?

A more affectionate position, despite its character of self-preservation, holds that we cannot afford (the pun is foreboding) to allow others' suffering to enter into our personal ken because its enormity will paralyze us or render us hopeless. There is again a strand of self-deception here but this version at least acknowledges that not all is well. The upshot, however, of each of these responses is that we should do nothing or we can do nothing to resolve either the existence or the meaning of collective unrequited suffering.

One further and treacherous rationalization abides in the mind of our own self-styled mandarin class, planet-wide. I refer here to a legitimation of human history by virtue of our collective monuments, historical moments of dazzling heroism, reflective accomplishment, and human creations greater than the sun, the moon, and the earth, as John Keats once wrote. Just how does one reconcile the magnificence of the Pyramids, Angkor Wat, and the Great Wall with how they were built, by whom, and at what human price. Can we afford, spiritually, psychologically, to reflect on the plight of our Irish and Chinese forbears as we Amtrak our way from Chicago to Los Angeles? Do we not then invoke the Spartacus syndrome forgetting the history of Thrace while thrilling to the courageous if abortive rage of one who carries for us just a glimmer of liberating possibility? I have no way out of this masking of the terror of our past. It would seem that Hegel was accurate when he described human history as a slaughter bench, riven with victims, rescued only in meaning by those "heroes" who epochally embody the *Phenomenologie des Geistes*. And just what do we think, only recently, of the contrast between our being appalled at the potential destruction of the archaeological treasures of ancient Mesopotamia and the also present destruction of the lives of thousands of children, as innocent in that debacle as they have been in all of the pockmarked millenia preceding? Is it not true that deep within us, the cultural remains take precedence, for, after all, they are immortal, whereas the children, like we, are mortal? And so it goes.

THE JOURNEY· AMELIORATION AS NECTAR

Some decades ago, an unusual refrain was heard over and over as part of a political campaign. After a litany of problems and afflictions, Robert Kennedy would say that "we can do better." This is hardly the stuff of rhetorical flourish, and, yet, the use of the word 'better' is a very important choice, for it replaces all of those halcyon words; cure, resolution, and those metaphors of comfort, as in to straighten out things, make everything whole, all on the way to a great society and a new world order. Unfortunately, these are the seeds of cynicism, for as I look over the wreckage of the human historical past, I see no hope for any resolution of anything humanly important.

This baleful perspective does not, however, obviate other responses such as healing, fixing *en passant,* rescuing, and yes, making, doing, and having things better. These approaches are actions on behalf of metaphysical amelioration, which holds that finite creatures will *always* be up against it and the best that we can do is to do better.

Yes, I acknowledge that the strategy of amelioration is vacant of the ferocious energizing that comes with commitment to an absolute cause, ever justifiable for some, somewhere, in spite of the nefarious results that most often accompany such political, religious, and social self-righteousness. A moral version of the maxim of Camus, cited above, would read, can I believe in helping when, *sub specie aeternitatis,* I hold that there is no ultimate resolution. Put differently, the original meaning of the ancient medical maxim, *primum non nocere,* was to do no harm. How and why did the maxim come to mean, keep the patient alive, at all cost, including the cost of dignity? What is it about us that cannot abide the sacrament of the moment as we reach for a solution, an end game, an explanation, a cure, nay, immortality?

I try as hard as I can to believe that the nectar is in the journey and not in its final destination. I stand with T.S. Eliot, who warns that "For us, there is only the trying. The rest is not our business." Perhaps I can describe my philosophical position as a Stoicism without foundation. Walt Whitman says it for me better than I can say it for myself. "The press of my foot to the earth springs a hundred affections, they scorn the best I do to relate them."

For what it is worth, and that, too, is a perilous question, I now believe, shakily, insecurely and barely, that life is worth living!

JOHN J. MCDERMOTT

Texas A & M University

QUESTIONING THE QUESTION OF THE WORTH OF LIFE*

Camus introduces *The Myth of Sisyphus*[1] with the claim that

> There is but one truly serious philosophical problem, and that is suicide. Judging whether life is or is not worth living amounts to answering the fundamental question of philosophy. All the rest—whether or not the world has three dimensions, whether the mind has nine or twelve categories—comes afterwards. These are games; one must first answer (*ibid.*, p. 3).

Should Camus be right and asking whether life is or is not worth living amounts indeed to asking the fundamental question of philosophy, then I have frittered away most of my philosophical life, playing what he dismisses as games, worrying about such issues as Heidegger's political engagement, or the relationship of freedom to truth, or the role of style in interpretation. How timely then the American Philosophical Association's call on my philosophical conscience in the form of an invitation to address the question: "What makes life worth living?"

But this was hardly how I felt. My first impulse was to decline. This had been a bad semester: unfinished essays, unanswered letters heaped on my desk; my wife facing a second cancer operation. "What makes life worth living?" Questions of life and death, thoughts of the worth of life, were very much on my mind, but not at all as philosophical questions—which may say something about philosophy, or at least about the kind of philosophy that has taken much of my time.

The invitation did make me think of my own life and of the lives of those close to me: of my mother, now 91, hoping to live long enough to finish the story of her turbulent life on which she is now working. Frail and not afraid to die, she still finds life most definitely worth living. I thought of my father, who died of a painful bone cancer, reached a point where he wanted to die and, had a revolver been handy, would have shot himself. He had loved life, but there did come a point where he no longer could find it worth living. Is life worth living? Whose life? Under what circumstances? In whose eyes? Are we the best judges of the worth of our lives?

The invitation came, as I said, at a bad time. I felt tired, not at all

---

* To be read in an APA symposium on What Makes Life Worth Living? December 30. John J. McDermott is co-symposiast, and Jay Rosenberg will comment. See this JOURNAL, this issue, 677–83, for McDermott's contribution.

[1] In *The Myth of Sisyphus and Other Essays*, Justin O'Brien trans (New York: Knopf, 1955).

0022-362X/91/8811/684–90

sure whether I wanted to invest the kind of effort that might make this into a worthwhile contribution.

I

Was it just intellectual laziness that made me feel that way? Or perhaps, worse, a refusal to live in that lucidity that, if Camus is right, is the reward of truly facing up to what he calls the fundamental question? Perhaps a fear of what such lucidity might reveal: the lack of meaning of a world riddled with suffering and dreadful accidents? A fear that I might have to recognize the truth of what Schopenhauer so tirelessly proclaimed: that pain in the end will always outweigh pleasure; that nothing can make life worth living?

The fact that I am standing here suggests disagreement with such sad wisdom, but my disagreement is not based on a reasoned judgment, indeed hardly based on a judgment at all; it expresses rather a stance toward my life, and beyond that toward life, a mood that governs me more than I have chosen it. Along with that mood goes the conviction that life should be such that it can be affirmed, that suicide should be unnecessary.

Can this conviction be supported with good reasons? It certainly cannot ground the judgment: "life as such is worth living." Quite the opposite: to be convinced that suicide should be unnecessary implies a belief that circumstances should not, but may be such that a rational actor will choose suicide. As Hannah Arendt[2] pointed out, "we know today that murder is not at all the worst thing one human being can do unto another, and that on the other hand, death is by no means what human beings fear most" (*ibid.*, p. 20). We fear unbearable pain, both physical and psychological, more than we do death and there are times when the thought of death alone makes bearable what would otherwise be unbearable. Thoughts of the possibility of suicide may thus console us. Schopenhauer[3] likens them to "the costly phial of poison to be found among the magnificent ornaments and apparel of oriental despots"; suicide is "for the case where the sufferings of the body, incapable of being philosophized away by any principles and syllogisms, are paramount and incurable" (*ibid.*, Vol. I, p. 91). Suicide here does not mean that life as such is judged not to be worth living, only that under certain circumstances it ceases to be worth living. Most of us have little difficulty imagining such situations. Such a belief does not at all conflict with the conviction that we should do everything to ensure that life remains and becomes ever more worth living.

[2] "Einleitung," in Hermann Broch, *Dichten und Erkennen: Essays*, vol. 1 (Zürich: Rhein, 1955).

[3] *The World as Will and Representation*, E. F. J. Payne, trans. (New York: Dover, 1969).

II

But I am supposed to address philosophers and philosophy would seem to demand more than personal recollections and commonplace declarations. Still, what is there to be said in answer to the question: What makes life worth living? Is life worth living? Is there something like a fact of the matter that determines the truth value of this proposition? Where would we look for that fact? From what standpoint could the answer be given? Is any positive answer in the end more than the expression of a quite blind, but difficult to eradicate will to live? To quote Schopenhauer once more:

> The will, as the metaphysical, is everywhere the boundary stone of every investigation beyond which this cannot go anywhere. From the original and unconditioned nature of the will . . . it is easy to explain that man loves above everything else an existence which is full of want, misery, trouble, pain, anxiety, and then again full of boredom, and which, were it pondered over and considered purely objectively, he would of necessity abhor; and that he fears above everything else the end of his existence, which is nevertheless for him the one and only thing certain. Accordingly, we often see a miserable figure, deformed and bent with age, want, and disease, appeal to us from the bottom of his heart for help for the prolongation of an existence, whose end would necessarily appear as altogether desirable, if it were an objective judgment that were the determining factor (*op. cit.*, Vol. II, p. 359).

Schopenhauer is right to insist that "the-will-to-live is not a consequence of the knowledge of life," but rather "that which is first and unconditioned" (*ibid.*, p. 360). Its fundamental importance should make us suspicious of all supposed arguments that this or that activity or possession makes life worth living as no more than special pleading, illusions the will to live casts ahead of itself. I think we have to grant Schopenhauer this much (and I shall have to return to this point): considered objectively, life has no worth—where much depends on understanding the nature and limits of such objective consideration. But we also must acknowledge how little such insight matters. As Schopenhauer knows full well, all the many considerations he has advanced to persuade his reader of the worthlessness of life, all the moving examples of pointless suffering he has assembled, are unlikely to gain a single convert to his pessimistic stance. Arguments for the worthlessness of life seem powerless before the will to live—a pointer to the impotence of pure reason.

Camus is not quite prepared to concede quite this much:

> Schopenhauer is often cited, as a fit subject for laughter, because he praised suicide [a misunderstanding of his position] while seated at a well-set table. This is no subject for joking. . . . In the face of such contradictions and obscurities must we conclude that there is no relationship between the opinion one has about life and the act one com-

> mits to leave it? Let us not exaggerate in this direction. In a man's attachment to life there is something stronger than all the ills in the world. The body's judgment is as good as the mind's, and the body shrinks from annihilation. We get into the habit of living before we get into the habit of thinking (*op. cit.*, pp. 7–8).

Not that the former habit is strong enough to prevent some people from dying "because they judge that life is not worth living" (*op. cit.*, p. 4). Does thought here place itself in opposition to the body and overrule its judgment?

I would like to question Camus's opposition of the body's judgment to the mind's. How are we to understand such a judging body? Does Camus perhaps attribute judgment to the body only to compensate for his own overly thin, too angelic, too Cartesian conception of the thinking subject? Could it be that the question "Is life worth living?"—this supposedly fundamental question of philosophy—rests on a questionable self-elevation of the spirit which leaves behind the embodied self and with it countless little pleasures that feed and are fed by our will to live, a self-elevation which despite its promise of a radical freedom is at bottom a self-alienation? In this connection, we should note the gap that separates the pleasures of everyday life, the many little meanings that are part of our usual dealings with persons and things, from that big, God-haunted meaning that in Camus's eyes could make life worth living:

> What I fail to understand is nonsense. The world is peopled with such irrationals. The world itself, whose single meaning I do not understand is but a vast irrational (*op. cit.*, p. 27).

But it is a false rationalism that opposes irrationality to meaning. Is it not precisely countless little irrationals that make life meaningful, whether it be an undeserved loving look, a melody, a flower-filled field. To be sure, I, too, recognize that I can live my life only as *one* life. But does this mean that *one* central meaning must preside over my life and gather it into a strongly integrated whole? I am suspicious of the sort of singlemindedness demanded by Kierkegaardian purity of heart or Heideggerian authenticity—it has to bring with it an inability to respond to the claims of the world; suspicious of a singlemindedness that leaps all too readily over the lesser problems that need to be addressed to make our lives better to the big question: Is life worth living?

III

Just as the question "Do you still love me?" is haunted by the (self-validating?) suspicion that love has passed,[4] so the question "Is life worth living?" is haunted by the (equally self-validating?) suspicion that our will to live may not at all mean that life is worth living, that the will to live is blind, that reason must call all claims to the worth of life into question as it objectively weighs its minuses and pluses. But can we even make sense of such an objective weighing?

---

[4] See Leszek Kolakowski, *The Presence of Myth*, Adam Czerniawski, trans. (Chicago. University Press, 1989), p. 48.

Initially at least, the question "Is life worth living?" seems easy enough to understand. The type seems familiar: we often ask ourselves when contemplating a certain action, say, one that involves risk or an expenditure of energy or funds, is it worth it? Is it worth, say, to undergo a certain operation? In reply, we might point to the chances that it will allow one to live longer or to the quality of life it will make possible. But such considerations presuppose that life is worth living. In such cases, there is a readily available context that allows us to place and address the question. That context furnishes us with relevant measures or criteria.

But when I question the worth of my life, no wider context offers itself which allows me to place and thus to get a handle on what is at issue. As I confront the possibility of my death, my life becomes a figure on an opaque ground of silence that gives greater weight to all that now pains me, but also to all that now claims and delights me. What makes life worth living? The knowledge that it is in my power to end it all invites me to weight its worth. But do we ever have the proper weights and the proper overview of our life, including all that might become of it, to give its correct weight?

To be sure, looking at a life from the outside, say, from a particular community's perspective, we may judge even a life that ended in suicide to have been very much worth living. Van Gogh, for example, must have reached a point when he no longer found his life worth living; but would we not call it, despite all his pain and suffering, despite whatever conclusion he may have arrived at, a worthwhile life? Were I to meet someone in his condition, I would try to convince him of this, get him to take himself and his present suffering a bit less seriously, to put his pain in perspective, lead him to recognize that it is our place in and our contribution to a larger community that makes life worthwhile. I might invite him to think of how artists and thinkers have helped to make countless lives more meaningful; or speak of those who by their kindness and love made others happier: Are their lives not justified and worthwhile?

I wonder whether such preaching would have much of an impact, for it presupposes just what is at issue: that life is worth living. Someone actually contemplating suicide is likely to remain unconvinced. And who is to call him wrong? I could offer him sentiments, but no good reasons to shake him out of his self-centered, self-destructive mood that lets him judge the kaleidoscope of life a meaningless jumble and life not worth living.

Should I then hand him *The Myth of Sisyphus,* which accepts the absurdity of our existence but insists that revolt

> . . . gives life its value. Spread out over the whole length of a life, it restores its majesty to that life. To a man devoid of blinders, there is no finer sight than that of an intelligence at grips with a reality that transcends it (*op. cit.,* p. 55).

The difficulty with such a proud nihilism that would replace "Belief in the meaning of life," which "always implies a scale of values, a choice, our preferences," with the knowledge that "one can live without appeal," with a pride that, in the face of the absurd, claims to discover the worth of life in lucidity and revolt, is that it has to render life without aims, render freedom irresponsible (*op. cit.,* p. 59), all choice weightless, a matter of indifference—even, finally, the choice of life or death. Camus proclaims that "Being aware of one's life, one's revolt, one's freedom, and to the maximum, is living, and to the maximum. Where lucidity dominates, the scale of values becomes useless" (*op. cit.,* p. 63). But where all scales of value becomes useless, choice becomes a matter of indifference as freedom degenerates into arbitrary spontaneity. With this life loses its worth.

IV

Earlier I suggested that we must grant Schopenhauer that, objectively considered, life has no worth. For what does it mean to consider something objectively? Objective consideration demands freedom from perspectival distortion. He who considers things objectively is not imprisoned in a particular here and now, but transcends all accidents of spatial and temporal location, transcends therefore also his embodied self. Kant thus distinguished subjective appearance, thought relative to the embodied, perceiving self, from the objects in themselves (not to be confused with things in themselves) thought relative to the transcendental subject, which, as the form of all possible experience, cannot be tied to any particular point of view. That these objects in themselves cannot be given in perception but must be reconstructed by us in thought should be evident. Such reconstruction is the task of science, which has its regulative ideal in the idea of a truly objective understanding of reality.

Such understanding requires precisely that self-elevation which is a presupposition of Camus's understanding of lucid thinking and of its opposition to what he calls the body's judgment. Objective consideration demands a reduction of reality that must elide, not just all that binds everyday experience to particular perspectives and points of view, but also the moods and interests that give reality worth or value. As being comes to be understood objectively as presence to a disembodied, disinterested consciousness, the world is transformed into meaningless pictures. As Schopenhauer points out, lost in this reduction is the significance of things, that "by virtue of which these pictures or images do not march past us, strange and meaningless, as they would otherwise inevitably do, but speak to us directly, are understood, and acquire an interest that engrosses our whole nature" (*op. cit.*, vol. I, p. 95). To the extent that we are committed to the objective consideration of reality, we shall have to agree with Wittgenstein's propositions in the *Tractatus logico-philosophicus*[5] that

> In the world everything is as it is and happens as it does happen. *In* it there is no value—and if there were, it would be of no value. . . . Hence also there can be no ethical propositions (*ibid.*, 6.41, 6.42).

To repeat, objectively considered, life has no worth. And if our understanding of what constitutes meaningful propositions is linked to an understanding of truth that requires such objectivity, then there can be no meaningful discourse of what makes life worth living. But this only invites careful and critical consideration of both the limits and the legitimacy of objective understanding.

V

In the *Philosophical Investigations,* Wittgenstein[6] writes "A philosophical problem has the form: 'I don't know my way about'" (*ibid.*, 123). To be sure, most problems having that form are not philosophical: losing one's way in a strange city, e.g., hardly suffices to make one a philosopher. I would suggest that this is so because our disorientation here is only superficial. We know what we must do to reorient ourselves; the problem poses itself against a background understanding that remains unquestioned. Philosophical problems have no such background. All philosophy is radical interpretation. The question shadowing all genuine philosophizing is: Where should I (we) be going? At the center of philosophy lies an ethical concern, born of the demand that human beings assume responsibility for their actions and the consequent refusal to rest content with maps and authorities that have come to be established, accepted, and taken for granted. So understood, philosophy remains alive only as long as the question "What is the right way" continues to be asked, because that way, and that means also what makes life worth living, remains questionable even as the idea of the right way haunts us. Such questioning assumes, in the face of all its obscurity, the worth of life. Both obscurity and worth manifest themselves in the countless ways in which we are called on, claimed, and moved by what we encounter. Such variety requires continuous and inescapably precarious interpretation of what is to be done.

KARSTEN HARRIES

Yale University

---

[5] C. K. Ogden, trans (New York: Routledge and Kegan Paul, 1933).

[6] G. E. M. Anscombe, trans. (New York: Macmillan, 1959).

NOTICE ABOUT 1992 APA EASTERN MEETING

The editors would like to announce that, because of our backlog problem and interest in publishing special issues, the Journal will no longer publish the program and symposium papers for the annual Eastern Division Meeting of the American Philosophical Association. Beginning in 1992, they will be published by the APA. For more information, please contact Shirley Anderson, Assistant Director, APA, University of Delaware, Newark, DE 19716.

NOTES AND NEWS

The Program in Ethics and the Professions at Harvard University invites applications for Fellowships in Ethics. Six fellowships will be awarded in 1992–93 to outstanding teachers and scholars who wish to develop their competence to address ethical issues in schools of business, government, law, medicine, and public policy. Fellows will participate in the weekly seminar of the program, attend courses in one of the professional schools or in the Graduate School of Arts and Sciences, and conduct their own research on ethics. The followship usually extends from September through July. Applicants usually hold a postgraduate degree in business, government/public policy, law, or medicine, or a doctorate in philosophy, political theory, or theology. Application deadline is January 6, 1992. For further information, please contact Jean McVeigh, Program in Ethics and the Professions, Harvard Univ., 79 Kennedy St., Cambridge, MA 02138.

The University Center for Human Values at Princeton University invites applications for Laurance S. Rockefeller Visiting Fellowships to be awarded for the academic year 1992–93 to outstanding teachers and scholars who are interested in devoting a year in residence at Princeton to writing about ethics and human values. Applicants are expected to have a doctorate or a professional postgraduate degree. The deadline for application materials is January 6, 1993, for fellowships beginning September 1992. For further information, please contact Valerie Kanka, Univ. Center for Human Values, 432 Robertson Hall, Princeton Univ., Princeton, NJ 08544.

CENTRAL LIBRARY

The City University of New York Graduate Center is happy to announce that Hartry Field will be joining the Program as Distinguished Professor and first holder of the John H. Kornblith Family Chair in the Philosophy of Science and Value.

The Department of Philosophy of the California State University/Chico is pleased to announce the appointments of Ellen Fox and Eric Gampel.

The Department of Philosophy of Denison University is pleased to announce the appointment of Eric C. Barnes as Assistant Professor.

The Department of Philosophy of Indiana University/Bloomington is pleased to announce the appointment of Michele Moody-Adams as Assistant Professor, and the promotions of Michael Morgan and Karen Hanson to Full Professors.

The Department of Philosophy at Occidental College is pleased to announce the appointment of Ann Cudd as Assistant Professor.

The Department of Philosophy at Ohio University is pleased to announce the appointments of Neil Tennant and Mark Wilson as Professors; William Taschek and Kathleen Cook as Assistant Professors; and Anna Greco as Instructor.

The Pomona College Philosophy Department is pleased to announce the appointments of Hilary Bok and Anthony Dardis as Assistant Professors; the promotion of Jay D. Atlas to Professor; and the accession of its chairman, Stephen A. Erickson, to the E. Wilson Lyon Professorship in the Humanities.

The Department of Philosophy at the University of California/Davis is pleased to announce the appointment of Jeff King as Assistant Professor.

The University of California/Riverside announces the appointments of Genoveva Marti and H. Carl Hoefer as Assistant Professors of Philosophy.

The Department of Philosophy of the University of Houston is pleased to announce the appointment of Anne Jacobson as Associate Professor of Philosophy.

The Department of Philosophy of the University of Illinois/Chicago is pleased to announce the appointment of Peter Hylton as Professor of Philosophy, effective Fall 1992.

The Department of Philosophy at Washington University/St. Louis is pleased to announce the appointments of Lawrence M. May as Professor, Marilyn A. Friedman as Associate Professor, and Hud Hudson as Assistant Professor.

From McGraw-Hill

FOR YOUR PHILOSOPHY COURSES

New for 1991

SOCIAL ETHICS
MORALITY AND SOCIAL POLICY, Fourth Edition
Thomas A. Mappes, *Frostburg State University*
Jane S. Zembaty, *University of Dayton*

QUESTIONS THAT MATTER
Third Edition
Ed. L. Miller, *University of Colorado — Boulder*

THEORIES OF HUMAN NATURE
CLASSIC AND CONTEMPORARY READINGS
Donald Abel, *St. Norbert College*

MORALITY AND THE GOOD LIFE
AN INTRODUCTION TO ETHICS THROUGH CLASSICAL SOURCES
Second Edition
Robert C. Solomon, *University of Texas — Austin*

For more information, please contact your local McGraw-Hill representative or write to McGraw-Hill College Division, Comp Processing & Control, P.O. Box 454, Hightstown, NJ 08520-0454.

THE JOURNAL OF PHILOSOPHY

is pleased to announce the publication of its

TWENTY-FIVE YEAR INDEX, 1964–1988

VOLUME LXI TO VOLUME LXXXV
AUTHORS, SUBJECTS, & REVIEWS

Supplement to its

FIFTY-YEAR INDEX, 1904–1953

TEN-YEAR SUPPLEMENT, 1954–1963

It is available for the rest of 1991 at special prices

25-year (1964–1988) $25.00 cloth/$10.00 paper

also available

50-year (1904–1953) $12.00 cloth 10-year (1954–1963) $3.50 paper

Shipping $5.00 (first)/$2.50 (each addl.)

Please send a check or money order to

Journal of Philosophy
704 Philosophy Hall
Columbia University
New York, New York 10027

Prices are in effect through December 31, 1991. Orders and payment must be received by then.

HARVARD

BOOTHS 303-305

MICHEL FOUCAULT
Didier Eribon
Translated by Betsy Wing
31 halftones/$27.95

A PROGRESS OF SENTIMENTS
Reflections on Hume's *Treatise*
Annette Baier
$39.95

WE THE PEOPLE
Volume 1: Foundations
Bruce Ackerman
The Belknap Press $24.95

LACAN
Malcolm Bowie
$19.95

ARISTOTLE
History of Animals, Books VII-X
Edited and Translated by D.M. Balme
Loeb Classical Library, L439 $15.50

VARIETIES OF MORAL PERSONALITY
Owen Flanagan
$34.95

UNIFIED THEORIES OF COGNITION
Allen Newell
$39.95

Michael Dummett

FREGE
Philosophy of Mathematics
$34.95

THE LOGICAL BASIS OF METAPHYSICS
William James Lectures
$34.95

ORIGINS OF THE MODERN MIND
Three Stages in the Evolution of Culture and Cognition
Merlin Donald
$27.95 cloth

Forthcoming in Spring '92

KANT AND THE EXACT SCIENCES
Michael Friedman

New in Paperback

TOWARD A FEMINIST THEORY OF THE STATE
Catharine A. MacKinnon
$12.95

SCHOPENHAUER AND THE WILD YEARS OF PHILOSOPHY
Rüdiger Safranski
Translated by Ewald Osers
$14.95

INNOCENCE AND EXPERIENCE
Stuart Hampshire
$9.95

LIBERALISM AND THE MORAL LIFE
Edited by Nancy L. Rosenblum
$14.95

SELECTIONS FROM CULTURAL WRITINGS
Antonio Gramsci
$16.95

Harvard University Press
79 Garden St., Cambridge, MA 02138

KLUWER ACADEMIC PUBLISHERS

including the D. Reidel Publishing Company
and Martinus Nijhoff Publishers imprints

Publishers of leading journals in:

Philosophy of Science
Logic
Phenomenology
History of Philosophy

Epistemology
Philosophy of Medicine
Philosophy of Law
etc.

Our journals include:

Studies in Philosophy and Education. Editor: David Ericson
Human Studies. Editor: George Psathas
Journal of the History of Biology. Editor: Everett Mendelsohn and Shirley A. Roe
Man and World. Editors: John Anderson, Joseph Kockelmans and Calvin Schrag
Philosophical Studies. Editor-in-chief: Keith Lehrer
Synthese. Editor-in-chief: Jaakko Hintikka
Theory and Decision. Editor-in-chief: W. Leinfellner
Argumentation. Editor-in-chief: Michel Meyer
Biology & Philosophy. Editor: Michael Ruse
Husserl Studies. Editors: J.N. Mohanty and Karl Schuhmann
Law and Philosophy. Editor: Alan Mabe
The Journal of Medicine and Philosophy.
Editors: H. Tristram Engelhardt, Jr. and Stuart Spicker
The Journal of Value Inquiry. Managing Editor: Robert Ginsberg
Theory and Society. Editor: Janet Gouldner
The Journal of Business Ethics. Editor-in-Chief: Alex Michalos
International Journal for Philosophy of Religion. Editor: Eugene Long and Frank R. Harrison III
Minds and Machines. Editor: James Fetzer
Linguistics and Philosophy. Editor: David Dowty

For more information on our publishing programs, contact:
Kluwer Academic Publishers
101 Philip Drive • Norwell, MA 02061 • (617) 871-6600

See our journals and books at APA Booths #208-210

Time, Freedom, and the Common Good

An Essay in Public Philosophy

Charles Sherover

"Charles M. Sherover's Time, Freedom, and the Common Good *is, to my knowledge, the most important philosophical articulation and defense of a democratic, capitalist republicanism since Hegel's* Philosophy of Right, *to which it is similar. With exemplary clarity and scholarship throughout, Sherover moves from phenomenological and pragmatic accounts of human temporality, freedom, and community to concrete political proposals which are eminently practical, relevant, and challenging. This book cannot be ignored."*

— Gregory R. Johnson
Critical Review

"Two features of this book deserve praise. The first is the interweaving of U. S. and European contemporary philosophical traditions. Sherover tries to conjugate phenomenology and pragmatism by emphasizing that democratic pluralism can gain more from a complex rendering of the experience of freedom in human life than from accounts of human motivations based on utilitarian considerations or rational choice theory.

"The second (praiseworthy) feature of Sherover's book is his attempt to overcome the, by now, traditional and slightly stale dichotomies pervasive in the philosophical and political debate in the United States. By this I mean the dichotomies between individual rights and the common good, the market and "big government," and "abstract foundationalism" and communitarian contextualism."

— Daniela Gobetti
American Political Science Review

State University of New York Press
c/o CUP Services • PO Box 6525
Ithaca, NY 14851 • 1-800-688-2877 (FAX orders)

Princeton

Valuing Life

John Kleinig

Abortion, euthanasia, capital punishment, war, genetic engineering and fetal experimentation, environmental and animal rights—these topics inspire some of today's most heated public controversies. And it is fashionable to pursue these debates in terms of the query "Under what conditions may life be disregarded or terminated?"

John Kleinig asks different questions: What may be said in behalf of life? Is livingness as such to be affirmed and respected? Is there an ascending order of plant, animal, and human life? Does human life possess a distinctive claim, or must we discriminate between humans that do and humans that do not have claims on us?

"Kleinig has written the most detailed, sustained analysis of the idea of life's value and of the arguments supporting that idea that I have seen. *Valuing Life* brings together in one place all the arguments and considerations relevent to making up one's mind about this part of ethical theory."—*James Rachels, University of Alabama at Birmingham*

Studies in Moral, Political, and Legal Philosophy
Marshall Cohen, Editor
Cloth: $35.00 ISBN 0-691-07388-0

New in paperback

Moral Tradition and Individuality

John Kekes

John Kekes develops the view that good lives depend on maintaining a balance between one's moral tradition and individuality. Our moral tradition provides the forms of good lives and the permissible ways of trying to achieve them. But to do so, he argues, we must grow in self-knowledge and self-control to make our characters suitable for realizing our aspirations. In addressing general readers as well as scholars, Kekes makes these philosophical views concrete by drawing on a rich variety of literary sources.

"[John Kekes] opens new vistas and his book is a delight to read. His theses make one itch to rethink the problems, but this is the mark of philosophical success."
—*Leslie Armour, Canadian Philosophical Reviews*
Now in paper: 14.95 ISBN 0-691-02348-4

Princeton University Press

41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 258-4900
ORDERS: 800-PRS-ISBN (777-4726) • OR FROM YOUR LOCAL BOOKSTORE

the edwin mellen press

Martinich, A. P. and Michael White
CERTAINTY AND SURFACE IN EPISTEMOLOGY AND PHILOSOPHICAL METHOD
Essays in Honor of Avrum Stroll
A collection of articles by distinguished philosophers from the US and Europe on two central topics in epistemology, certainty and surfaces. Four of the ten articles discuss Avrum Stroll's Surfaces (1988).
0-7734-9711-0 [PCP 32*] $59.95

Aarons, Victoria and Willis Salomon, editors
RHETORIC AND ETHICS
Historical and Theoretical Perspectives
Addresses the classical connection between rhetoric and ethics, its fragmentation during the 17th century, and the reunification of rhetoric and ethics in the 20th century.
0-88946-212-7 [PCP 27*] $59.95

Basinger, David and Randall
PHILOSOPHY AND MIRACLE
The Contemporary Debate
Presents an assessment of recent discussions of the concept of miracle that have taken place within the analytic tradition.
0-88946-327-1 [PCP 2*] $49.95

mellen

See our Acquisitions Editor in *Booth 212* at the APA.

The Edwin Mellen Press invites manuscript proposals for scholarly works in philosophy and other fields.

The purpose of the Edwin Mellen Press is to promote and publish scholarship -- and "contribution to scholarship" is its sole criterion for accepting a manuscript for publication.

P.O. Box 450 Lewiston, NY 14092

philosophy

MACMILLAN ANNOUNCES A NEW VOLUME
IN *PHILOSOPHICAL TOPICS*

IMMORTALITY

Edited with an introduction
and annotated bibliography by

PAUL EDWARDS

Thirty-four selections covering the mind-body problem, personal identity as well as Eastern and Western views on immortality. Believers and skeptics are equally represented. Reincarnation is covered extensively in both the readings and the introduction.

Copies available December 1991

MacMillan Publishing Company, College Division
445 Hutchinson Avenue, Columbus Ohio 43235
Attention: Fred Hamden, Marketing Manager

CALIFORNIA

BOOI II I I 9

NEW IN HARDCOVER

Ernst Bloch
Heritage of Our Times
Translated with Introduction, by Neville & Stephen Plaice
$40 00 cloth *Weimar and Now*

Troyen A. Brennan, M D , J D.
Just Doctoring: Medical Ethics in the Liberal State
$34 95 cloth

Norman O. Brown
Apocalypse and/or Metamorphosis
$22 50 cloth

Carnes Lord & David O'Connor, Editors
Essays on the Foundations of Aristotelian Political Science
$39 95 cloth

Nishitani Keiji
Nishida Kitaro
$25 00 cloth *Nanzan Studies in Religion and Culture*

David L. Norton
Democracy and Moral Development: A Politics of Virtue
$29 95 cloth

K.S. Shrader-Frechette
Risk and Rationality: Philosophical Foundations for Populist Reforms
$15 95 paper, $39 95 cloth

NEW IN PAPERBACK

Brian Barry
Theories of Justice
California Series on Social Choice and Political Economy $14.95 paper

Alan H. Goldman
Empirical Knowledge
$14.95 paper

Immanuel Kant
Observations on the Feeling of the Beautiful and the Sublime
Translated & Edited by John T Goldthwait
$10 95 paper

Nel Noddings
Women and Evil
$12 95 paper

Hans Reichenbach
The Direction of Time
Edited by Maria Reichenbach
New by Introduction Hilary Putnam
$14 95 paper

G S. Rousseau, Editor
The Languages of Psyche: Mind and Body in Enlightenment Thought
$16 95 paper, $55 00 cloth
Publications from the Clark Library Professorship

James J. Sheehan & Morton Sosna, Editors
The Boundaries of Humanity: Humans, Animals, Machines
$12 95 paper, $35 00 cloth

University of California Press
Berkeley Los Angeles New York Oxford

•New for 1992•

CRITICAL THINKING,
Third Edition

Brooke Noel Moore and
Richard Parker
California State University,
Chico

This new edition builds on the features that have made ***Critical Thinking*** the best-selling critical thinking text available. Of special importance, ***Critical Thinking*** now includes a separate chapter on writing argumentative essays and new, easily gradable writing improvement exercises. Other new features include guidelines for recognizing and avoiding racist and sexist language, and two chapters, by Anita Silvers and Sterling Harwood, covering moral, legal, and aesthetic reasoning.

BEYOND THE WESTERN TRADITION: ***Readings in Moral and Political Philosophy***

Daniel Bonevac, William Boon,
and Stephen Phillips
The University of Texas at
Austin

A supplemental collection of the most influential readings in African, Asian, and Southern Mediterranean moral and political philosophy, ***Beyond the Western Tradition*** presents both classic and contemporary readings for each tradition.

For more information about any Mayfield text after the meeting, contact your Mayfield sales representative or call the sales department at 1(800)433-1279.

TODAY'S MORAL ISSUES: ***Classic and Contemporary Perspectives***

Daniel Bonevac
The University of Texas at Austin

This anthology of 63 readings draws from both classic and contemporary readings to explore today's most pressing moral issues, including censorship, abortion, euthanasia, and capital punishment. The classic readings provide important theoretical background and give students a sense of the historical and philosophical contexts of the issues. The authors of the contemporary readings are prominent in public debate over these issues. Some are philosophers, but selections by economists, politicians, environmental scientists, sociologists, lawyers, journalists, and essayists are also included.

THINKING CRITICALLY ABOUT ETHICAL ISSUES,
Third Edition

Vincent Ryan Ruggiero, Emeritus
State University of New York at
Delhi

This is a lively, highly accessible introduction to the role of critical thinking in moral reasoning. The emphasis is on the *application* of ethical principles to the kinds of moral issues students face in their daily lives. More than 350 brief, thought-provoking case studies are provided for analysis and discussion.

Please stop by and see us at booths 410 and 412.

Mayfield

1240 VILLA STREET, MOUNTAIN VIEW, CA 94041

Studies in Continental Thought

The Principle of Reason
Martin Heidegger
Translated by Reginald Lilly
cloth $27.50

Sartre's Political Theory
William L. McBride
cloth $37.50 • paper $12.95

Narrative and the Self
Anthony Paul Kerby
cloth $22.50

Re-reading Levinas
Ed. by Robert Bernasconi & Simon Critchley
cloth $39.95 • paper $14.95

The Indiana Series in the Philosophy of Technology

Instrumental Realism
The Interface between Philosophy of Science and Philosophy of Technology
Don Ihde
cloth $29.95 • paper $12.50

The Indiana Series in the Philosophy of Religion

Kant's Philosophy of Religion Reconsidered
Ed. by Philip J. Rossi & Michael J. Wreen
cloth $25.00

Writings of Charles S. Peirce
A Chronological Edition
Volume 5, 1884-1886
Editors of the Peirce Edition Project
cloth $65.00

Feminism and Philosophy
Perspectives on Difference and Equality
Moira Gatens
cloth $39.95 • paper $14.95

Revaluing French Feminism
Critical Essays on Difference, Agency, and Culture
Ed. by Nancy Fraser & Sandra Lee Bartky
cloth $35.00 • paper $12.95

The Jewish Bible after the Holocaust
A Re-reading
Emil L. Fackenheim
cloth $27.50

Universe of the Mind
A Semiotic Theory of Culture
Yuri Lotman
Translated by Ann Shukman
cloth $45.00

The Sense of Change
Language as History
Michael Shapiro
Advances in Semiotics
cloth $29.95

Contending Rhetorics
Writing in Academic Disciplines
George L. Dillon
cloth $27.50

Legal Theory, Political Theory, and Deconstruction
Against Rhadamanthus
Matthew H. Kramer
cloth $35.00

Politics/Power/Culture
Postmodernity and Feminist Political Theory
Ed. by Kathy E. Ferguson & Kirstie M. McClure
A Special Issue of DIFFERENCES,
Vol. 3, No. 1 paper $12.95

Role Playing and Identity
The Limits of Theatre as Metaphor
Bruce Wilshire
Now in paperback! $14.95

Time, Narrative, and History
David Carr
Now in paperback! $10.95

Social Epistemology
Steve Fuller
Science, Technology, and Society
Now in paperback! $12.95

BOOTHS
302 & 304

INDIANA
University Press
Bloomington, IN 47404
1-800-842-6796

MINNESOTA PHILOSOPHY

GIORGIO AGAMBEN

LANGUAGE AND DEATH

The Place of Negativity

Explores the symbiosis of philosophy and literature.

$14.95 paper

GEORGES VAN DEN ABBEELE

TRAVEL AS METAPHOR

From Montaigne to Rousseau

Underscores the foundational force of travel in early modern French philosophy.

$14 95 paper

GUNNAR OLSSON

LINES OF POWER/ LIMITS OF LANGUAGE

An innovative social-theoretical language that explodes the very concept of disciplinary boundaries.

$19.95 cloth

J. CLAUDE EVANS

STRATEGIES OF DECONSTRUCTION

Derrida and the Myth of the Voice

The first detailed critical study of Derrida's interpretation and critique of Husserl.

$14.95 paper

RAYMOND S. FRANKLIN

SHADOWS OF RACE AND CLASS

Makes new connections between race and class and a proposal for radical urban social change.

$14.95 paper

MICHAEL J. SHAPIRO

READING THE POSTMODERN POLITY

Political Theory as Textual Practice

Demonstrates how textuality is inherent to political practice.

$15.95 paper

PAUL ZUMTHOR

TOWARD A MEDIEVAL POETICS

Integrates philosophy, history, and sociology to establish the dynamics of textual production in the Middle Ages.

$44.95 cloth

FRIEDRICH SCHLEGEL

PHILOSOPHICAL FRAGMENTS

Makes available in English the work of a man whom Rene Wellek has called "one of the greatest critics of history."

$11 95 paper

UNIVERSITY OF MINNESOTA PRESS

Visit booths 102 & 104 at the Eastern Division APA convention for a 20% discount on these and other titles. Or phone 1-800-388-3863.

KEY CONTEMPORARY THINKERS

Donald Davidson

Simon Evnine. "I had not dared to hope such a clear and sensible book would, or even could, be written about my work. . . . I can't agree with everything in the book. . . . This is all to the good: a book as interesting and intelligently sympathetic could not have been written by someone without convictions of his own."—Donald Davidson. Cloth, $32.50; paper, $10.95

Robert Nozick

PROPERTY, JUSTICE, AND THE MINIMAL STATE

Jonathan Wolff. "A serious and substantial investigation into the political philosophical questions that Nozick's work raises."—Arthur Kuflik, University of Vermont. Cloth, $32.50; paper, $9.95

STANFORD SERIES IN PHILOSOPHY

Has Semantics Rested on a Mistake?

AND OTHER ESSAYS

Howard Wettstein. These tightly focused essays record how a long-term study of Gottlob Frege has persuaded the author that Frege's pivotal distinctions between sense and reference, and his attendant philosophical views about language and thought, are unsatisfactory. $29.50

Nietzsche: The Body and Culture

PHILOSOPHY AS A PHILOLOGICAL GENEALOGY

Eric Blondel. Translated by Seán Hand. "This is a very exciting book. . . . The notions Blondel puts at the center of his study . . . are central to the whole set of concerns that make Nietzsche so important."—Bernard Williams, University of Cambridge. Cloth, $39.50; paper, $14.95

The Political Ontology of Martin Heidegger

Pierre Bourdieu. "The single most illuminating contribution to an understanding of the social and political meaning of Heidegger's work. . . . The book stands apart . . . by virtue of the originality, clarity, and soundness of its point of view."—Jean-Joseph Goux, Brown University. $22.50

Confrontations

DERRIDA / HEIDEGGER / NIETZSCHE

Ernst Behler. Translated, with an Afterword, by Steven Taubeneck. Because a confrontation with Heidegger is central to Derrida's writings (and because this confrontation is always bound up with Nietzsche), this book not only gives an account of the "New Nietzsche," but also provides a genealogy of Derrida's own writings. Cloth, $35.00; paper, $11.95

Chronotypes

THE CONSTRUCTION OF TIME

Edited by John Bender and David E. Wellbery. Philosophers, historians, literary critics, sociologists, and anthropologists examine the ways individuals, societies, and cultures make sense of time by constructing it in diverse patterns. Cloth, $37.50; paper, $12.95

Stanford University Press
STANFORD, CA 94305

TITLES FROM LANG

The Co-Existence of God and Evil
by Jane Mary Trau
1991 111 pp. 0-8204-1380-1 $27.95

"New thinking concerning the problem of evil is always to be welcome. Professor Trau's book...adds a new and intriguing element to the continuing discussion." —John Hick, Claremont Graduate School

Matter, Morals and Medicine
The Ancient Greek Origins of Science, Ethics and the Medical Profession
by Michael Jerome Carella
1991 361 pp 0-8204-1432-8
hardback $43.95 pb $23.95

How Judges Reason
by Joel Levin
1991 275 pp 0-8204-1549-9 TBA

Peirce and Law
Issues in Pragmatism, Legal Realism and Semiotics
Edited by Roberta Kevelson
Semiotics and the Human Sciences, Vol. 1
1991 225 pp 0-8204-1519-7 $43.95

For What May I Hope?
Thinking with Kant and Kierkegaard
by Gene Fendt
1990 224 pp. 0-8204-1387-9 $44.95

Philosophy of the Gita
by Ramesh N. Patel
1991 311 pp 0-8204-1416-6 $47.95

"Philosophy of the Gita *gives a fresh, new interpretation of the* Gita, *it presents a critical discussion of earlier translations, and it offers a novel philosophical background for the* Gita."—Paul F. Schmidt, Chair, Emeritus, Department of Philosophy, University of New Mexico

Peter Lang Publishing

PARAGON ISSUES IN

WOMAN AND THE HISTORY OF PHILOSOPHY

NANCY TUANA,
University of Texas, Dallas

Using the works of Aristotle, Plato, Descartes, Rousseau, Kant, Hume, Locke, and Hegel, the author offers a detailed examination of the history of philosophical theories concerning woman's nature and traits viewed as feminine. Topics covered include: the second sex; women and reason; women and morality; private woman, public man; the future of "feminine" and "masculine."

"Indispensable for introducing undergraduates to gender issues in the history of philosophy."
—Susan Bordo, *LeMoyne College*

Paper, ISBN 1-55778-194-X, $14.95

CRITICAL THEORY

The Essential Readings

DAVID INGRAM,
Loyola University, Chicago
& JULIA SIMON-INGRAM,
Washington University, St. Louis

Selections from the most important works of the Critical Theorists—Theodor Adorno, Max Horkheimer, Herbert Marcuse, and Jürgen Habermas—followed by a critique of their work by prominent post-structuralists, post-modernists, and feminists. A companion volume to *Critical Theory and Philosophy.*

Paper, ISBN 1-55778-353-5, $15.95

PHILOSOPHY AND COGNITIVE SCIENCE

JAMES H. FETZER,
University of Minnesota

An introduction to one of the most exciting new fields of inquiry today, which includes computer science, artificial intelligence, and linguistic/cognitive psychology. From Descartes' *Cogito* to the connectionist conception, scientific and philosophical inquiry are used to explore our modern conceptions of language, mentality, and mind. May be used as a companion text to *Foundations of Cognitive Science* by Jay L. Garfield.

Paper, ISBN 1-55778-195-8, $13.95

FOUNDATIONS OF COGNITIVE SCIENCE

The Essential Readings

JAY L. GARFIELD,
Hampshire College, Amherst

A reader in the philosophical foundations of cognitive science, this volume of 25 essays includes Pylyshyn, Lucas, Fodor, Chomsky, and many others. Each section explains the principal questions addressed, their importance in the philosophical foundations of cognitive science, and the range of positions taken with respect to them. May be used as a companion reader to *Philosophy and Cognitive Science* by James H. Fetzer.

Paper, ISBN 1-55778-257-1, $17.95

AFRICAN PHILOSOPHY

The Essential Readings

TSENAY SEREQUEBERHAN,
Hampshire College, Amherst

In 1988 the American Philosophical Association officially recognized African philosophy, giving this area of research special status. This volume traces the origins and development of African philosophy—including ethno-philosophy, professional philosophy, philosophic sagacity, and nationalist-ideological philosophy.

Paper, ISBN 1-55778-309-8, $13.95

PHILOSOPHY... PARAGON ISSUES IN PHILOSOPHY

CRITICAL THEORY AND PHILOSOPHY

DAVID INGRAM,
Loyola University, Chicago

An introduction to four major representatives of the Frankfurt School: Theodor Adorno, Max Horkheimer, Herbert Marcuse, and Jürgen Habermas. Beginning with Kant, Hegel, and Marx, and ending with post-structuralism, post-modernism, and feminism, it is accessible to those with little background in one of the most influential intellectual movements of this century.

Paper, ISBN 1-55778-201-6, $12.95

METAPHYSICS

A Contemporary Introduction

JOHN F. POST, Vanderbilt University

This vigorously argued introduction brings an exciting new philosophical perspective to traditional metaphysics and updates the field.

Paper, ISBN 1-55778-204-0, $16.95

MEANING AND TRUTH

The Essential Readings in Modern Semantics

JAY L. GARFIELD,
Hampshire College, Amherst,
& MURRAY KITELEY,
Smith College

The first anthology comprising the full canon in the philosophy of language and linguistic semantics. Selections begin with the classic essay by Mill, *Of Names and Propositions*, include such standards as *General Semantics* by Lewis and *Actualism* by Platinga, and conclude with a chapter on *Linguistic Approaches to Semantics.*

Paper, ISBN 1-55778-300-4, $17.95

LIVING THE GOOD LIFE

An Introduction to Moral Philosophy

GORDON GRAHAM,
University of St. Andrews, Scotland

What is the best life a human being can lead? An introduction for students to the principal ideas in Western moral philosophy. While focusing on historically significant thinkers from the past, it encourages readers to formulate their own opinions.

Paper, ISBN 1-55778-235-0, $12.95

PHILOSOPHY OF SPORT

DREW A. HYLAND,
Trinity College, Hartford

Covers a wide range of practical and theoretical issues which touch on sociological, psychological and philosophical themes—including self-knowledge, heroism, risk-taking, sexism, the connection between mind and body, and ethical issues.

Paper, ISBN 1-55778-189-3, $15.95

ATTENTION PROFESSORS:
To obtain examination copies, send your request on school letterhead, indicating your name and academic position, course name and numbers, projected enrollment, and date your adoption decision is due.

PARAGON HOUSE
90 FIFTH AVENUE
NEW YORK, NY 10011
1-800-PARAGON

Princeton

NEW SELECTIONS

Bernard Faure

The Rhetoric of Immediacy

A Cultural Critique of Chan / Zen Buddhism

Cloth: $39.50

Thomas Gould

The Ancient Quarrel between Poetry and Philosophy

Cloth: $39.50

C. G. Jung

Translated by Beatrice M. Hinkle

With a new introduction by William McGuire

Psychology of the Unconscious

A Study of the Transformations and Symbolisms of the Libido

Supplementary Volume B to The Collected Works of C. G. *Jung*

Bollingen Series XX:B

Cloth: $29.95

Søren Kierkegaard

Edited and translated, with introduction and notes, by Howard V. Hong and Edna H. Hong

Practice in Christianity

Kierkegaard's Writings, XX

Paper: $14.95   Cloth: $49.50

John Kleinig

Valuing Life

Studies in Moral, Political, and Legal Philosophy

Marshall Cohen, Editor

Cloth: $35.00

Livia Kohn

Early Chinese Mysticism

Philosophy and Soteriology in the Taoist Tradition

Paper: $14.95   Cloth: $45.00

Edited by Karl H. Potter and Sibajiban Bhattacharyya

Encyclopedia of Indian Philosophies

Volume VI: Indian Philosophical Analysis: Nyāya-Vaiśeṣika from Gaṅgeśa to Raghunātha Śiromaṇi

Cloth: $65.00

Nicholas Rescher

A System of Pragmatic Idealism

Volume I: Human Knowledge in Idealistic Perspective

Cloth: $37.50

Mark Turner

Reading Minds

The Study of English in the Age of Cognitive Science

Cloth: $29.95

Jacob Viner

Edited by Douglas A. Irwin

Essays on the Intellectual History of Economics

Cloth: $49.50

Please visit us
at Booth 403

NEW IN PAPERBACK

F. M. Cornford
From Religion to Philosophy
A Study in the Origins of Western Speculation
Mythos Series
Paper: $12.95

Mircea Eliade
Translated from the French by Willard R. Trask
Yoga ✣
Immortality and Freedom
Bollingen Series LVI
Paper: $14.95

Mary Louise Gill
Aristotle on Substance
The Paradox of Unity
Paper: $14.95 Cloth: $35.00

John Kekes
Moral Tradition and Individuality
Paper: $14.95 Cloth: $32.50

Søren Kierkegaard
Edited and translated, with introduction and notes, by Howard V. Hong and Edna H. Hong
The Concept of Irony
With Continual Reference to Socrates
Together with Notes of Schelling's Berlin Lectures
Kierkegaard's Writings, II
Paper: $17.95 Cloth: $69.50

Søren Kierkegaard
Edited and translated, with introduction and notes, by Howard V. Hong and Edna H. Hong
For Self-Examination / Judge for Yourself!
Kierkegaard's Writings, XXI
Paper: $12.95 Cloth: $35.00

Richard Kraut
Aristotle on the Human Good
Paper: $16.95 Cloth: $42.50

Edited, with an introduction, by Rudolf A. Makkreel and Frithjof Rodi
Selected Works of Wilhelm Dilthey: Volume I
Introduction to the Human Sciences
Paper: $18.95 Cloth: $59.50

Alan Roland
In Search of Self in India and Japan
Toward a Cross-Cultural Psychology
Paper: $14.95 Cloth: $39.50

Robert A. F. Thurman
The Central Philosophy of Tibet
A Study and Translation of Jey Tsong and Khapa's *Essence of True Eloquence*
Princeton Library of Asian Translations
Paper: $18.95

Winner of the Present Tense *Joel H. Cavior Literary Award in History*
Yirmiyahu Yovel
Spinoza and Other Heretics
The Marrano of Reason
The Adventures of Immanence
Marrano: Paper: $12.95 Cloth: $29.95
Adventures: Paper: $12.95 Cloth: $35.00
Cloth Set: $54.00

Please visit us at Booth 403

Princeton University Press
41 WILLIAM ST. • PRINCETON, NJ 08540
ORDERS: 800-PRS-ISBN (777-4726)
OR FROM YOUR LOCAL BOOKSTORE

CHICAGO & BOOTH 411

PHENOMENOLOGY AND DECONSTRUCTION, VOLUME I

The Dream is Over

Robert Denoon Cumming

Paper $17 95 176 pages Cloth $49.95

THE DIALECTICAL NECESSITY OF MORALITY

An Analysis and Defense of
Alan Gewirth's Argument to the
Principle of Generic Consistency

Deryck Beyleveld

With a Foreword by Alan Gewirth

Paper $22 50 520 pages (est) Cloth $55 00

THEORIES OF TRANSLATION

An Anthology of Essays
from Dryden to Derrida

Edited by **John Biguenet**
and **Rainer Schulte**

Paper $12.95 312 pages Cloth $27.50

GOD WITHOUT BEING

Hors-Texte

Jean-Luc Marion

With a Foreword by David Tracy
Translated by Thomas A. Carlson

Cloth $32.00 264 pages
Religion and Postmodernism series

THE REASON OF FOLLOWING

Christology and the Ecstatic I

Robert P. Scharlemann

Cloth $32.50 224 pages
Religion and Postmodernism series

THE LANGUAGE OF FIRST-ORDER LOGIC

Including the Macintosh™ Program,
Tarski's World 3.0

Jon Barwise *and*
John Etchemendy

Paper $27.50 270 pages
includes 3.5-inch diskette
Distributed for the Center for the Study of
Language and Information

TRAGEDY – CONTRADICTION AND REPRESSION

Richard Kuhns

Cloth $27 50 200 pages

THE RHETORIC OF MORALITY AND PHILOSOPHY

Plato's *Gorgias* and *Phaedrus*

Seth Benardete

Cloth $29 95 216 pages

THE FOUCAULT EFFECT

Studies in Governmentality

Edited by **Graham Burchell,**
Colin Gordon, *and* **Peter Miller**

Paper $19 95 320 pages Cloth $58 00

INTERPRETATIONS OF CONFLICT

Ethics, Pacifism, and the Just-War
Tradition

Richard B. Miller

Paper $17 95 296 pages Cloth $48.00

CODE OF PEACE

Ethics and Security in the World
of the Warlord States

Dorothy V. Jones

Cloth $24 95 220 pages

THE NOBLE SAVAGE

Jean-Jacques Rousseau, 1754-1762

Maurice Cranston

Cloth $32 50 416 pages

And in Paper —

JEAN-JACQUES

The Early Life and Work of
Jean-Jacques Rousseau, 1712-1754

Paper $17.95 382 pages

CHICAGO ❦ BOOTH 411

NIETZSCHE AND ASIAN THOUGHT
Edited by **Graham Parkes**
Cloth $27.50 272 pages

CROSSINGS
Nietzsche and the Space of Tragedy
John Sallis
Paper $11 95 168 pages Cloth $24 95

NIETZSCHE'S NEW SEAS
Explorations in Philosophy, Aesthetics, and Politics
Edited by **Michael Allen Gillespie** *and* **Tracy B. Strong**
Paper $13 95 248 pages

MAIMONIDES' ETHICS
The Encounter of Philosophic and Religious Morality
Raymond L. Weiss
Cloth $29 95 232 pages

RABBINIC POLITICAL THEORY
Religion And Politics in the Mishnah
Jacob Neusner
Paper $22.50 284 pages Cloth $49 95
Chicago Studies in the History of Judaism

ISAIAH BERLIN
A Celebration
Edited by **Edna Ullmann-Margalit** *and* **Avishai Margalit**
Cloth $27 50 225 pages

Now in Paper

MARTIN HEIDEGGER
George Steiner
With a new Introduction
Paper $12 95 208 pages

JOHN LOCKE'S LIBERALISM
Ruth W. Grant
Paper $13 95 230 pages

HEGEL'S CRITIQUE OF LIBERALISM
Rights in Context
Steven B. Smith
Paper $14 95 272 pages

CUNNING INTELLIGENCE IN GREEK CULTURE AND SOCIETY
Marcel Detienne *and* **Jean-Pierre Vernant**
Translated by Janet Lloyd
Paper $16 95 344 pages

GIORDANO BRUNO AND THE HERMETIC TRADITION
Frances A. Yates
Paper $16.95 480 pages

THE UNIVERSITY OF CHICAGO PRESS
1891-1991 *5801 South Ellis Avenue Chicago, IL 60637*

The Dialogues of Plato

Volume II: The Symposium

Translated with Commentary by R.E. Allen

"A readable and interesting book that is likely to be the standard work on the *Symposium* for a long time to come."—Robert G. Turnbull $25.00

The Idea of Wilderness

From Prehistory to the Age of Ecology

Max Oelschlaeger

"An absolutely crucial and timely benchmark in ecophilosophic thought and history."—George Sessions $29.95

William of Ockham

Quodlibetal Questions

Volume 1 (I-IV) translated by **Alfred J. Freddoso** and **Francis E. Kelley**
Volume 2 (V-VII) translated by **Alfred J. Freddoso**

The first English translation of William of Ockham's *Quodlibetal Questions*, based on the recent critical edition of Ockham's theological and philosophical works. Two-volume set, $100.00

Yale Library of Medieval Philosophy

Benjamin Constant and the Post-Revolutionary Mind

Biancamaria Fontana $25.00

The Life and Thought of Friedrich Engels

A Reinterpretation

J.D. Hunley $22.50

Affirmative Action and Justice

A Philosophical and Constitutional Inquiry

Michel Rosenfeld $30.00

Yale

Plato's Dialectical Ethics

Phenomenological Interpretations Relating to the Philebus

Hans-Georg Gadamer; translated and with an introduction by **Robert M. Wallace**

This classic work by Hans-Georg Gadamer—now translated into English for the first time—offers an extensive and imaginative interpretation of Plato's *Philebus*. $27.50

Inconvenient Fictions

Literature and the Limits of Theory

Bernard Harrison

A philosopher explores the differences between humanism and deconstructionism in literary criticism. $35.00

Difference and Subjectivity

Dialogue and Personal Identity

Francis Jacques; translated by **Andrew Rothwell**

"This new theory of subjectivity... addresses one of the most important concerns of modern philosophy."—Georg Mohr, *Revue de theologie et philosophie* $40.00

1991

On Justice

An Essay in Jewish Philosophy
L.E. Goodman

"A remarkable statement of what we all have to learn from the Jewish tradition of thought and practice." —Alasdair MacIntyre $37.50

Psychoanalysis and Ethics

Ernest Wallwork

A pathbreaking philosophical examination of the ethical relevance of psychoanalysis. $35.00

The Origins of Aristotelian Science

Michael Ferejohn $24.50

Philosophical Hermeneutics and Literary Theory

Joel Weinsheimer $19.50

The Self in Social Theory

A Psychoanalytic Account of Its Construction in Plato, Hobbes, Locke, Rawls, and Rousseau
C. Fred Alford $30.00

Architectural Principles in the Age of Historicism

Robert Jan van Pelt and Carroll William Westfall $35.00

Now available in paperback

Philosophy in World Perspective

A Comparative Hermeneutic of the Major Theories

David A. Dilworth

"Opens new interpretive possibilities for intra- and inter-textual reflection."—Edith Wyschogrod $14.00

The Ancients and the Moderns

Rethinking Modernity

Stanley Rosen

"The book...will make a significant difference in the way we view philosophy over the coming decades."—Kenley Dove $12.95

An Interpretation of Religion

Human Responses to the Transcendent

John Hick

Winner of the 1991 Louisville Grawemeyer Award $18.95

Democracy and Its Critics

Robert A. Dahl

"A work of extraordinary intelligence and...wisdom."—Robert N. Bellah, *N.Y. Times Book Review* $17.00

The Nature of Politics

Roger D. Masters $12.95

See these and other new titles at Booth 414.

Yale University Press
92A Yale Station
New Haven, CT 06520

Philosophy from Cambridge

Consciousness and Revolution in Soviet Philosophy
From the Bolsheviks to Evald Ilyenkov
David Bakhurst

Mind and Common Sense
Philosophical Essays on Common Sense Psychology
Radu J. Bogdan, Editor

The Cambridge History of Medieval Political Thought, c.350–c.1450
J. H. Burns, Editor

The Cambridge History of Political Thought 1450–1700
J. H. Burns, Editor
With Mark Goldie

The Semantic Tradition from Kant to Carnap
To the Vienna Station
J. Alberto Coffa
Linda Wessels, Editor

The Philosophical Writings of Descartes
Volume III:
The Correspondence
Rene Descartes
John Cottingham et al, Editors

Trials and Punishments
R. A. Duff

Interpersonal Comparisons of Well-Being
Jon Elster and John E. Roemer, Editors

Psychology
Stephen Everson, Editor

Liability and Responsibility
Essays in Law and Morals
R. G. Frey and Christopher W. Morris, Editors

Violence, Terrorism, and Justice
R. G. Frey and Christopher W. Morris, Editors

On the Nature and Existence of God
Richard Gale

Liberal Purposes
Goods, Virtues, and Diversity in the Liberal State
William A. Galston

Spectacles and Predicaments
Essays in Social Theory
Ernest Gellner

The Future of Folk Psychology
Intentionality and Cognitive Science
John D. Greenwood, Editor

Wittgenstein
Centenary Essays
Ludwig J. Wittgenstein
A. Phillips Griffiths, Editor

Plausible Worlds
Geoffrey Hawthorn

The Philosophy of Quantum Mechanics
Richard A. Healey

Sensations
A Defense of Type Materialism
Christopher S. Hill

Autonomy and Self-Respect
Thomas E. Hill, Jr.

Saussure
Signs, System and Arbitrariness
David Holdcroft

A Morally Deep World
An Essay on Moral Significance and Environmental Ethics
Lawrence E. Johnson, Editor

Philosophy from Cambridge

Philosophy and Politics
G. M. K. Hunt, Editor

Moral Legislation
A Legal-Political Model for Indirect Consequentialist Reasoning
Conrad D. Johnson

The Metaphysics of Morals
Immanuel Kant
Mary Gregor, Translator

The Language of Art History
Salim Kemal and Ivan Gaskell, Editors

Explanation and its Limits
Dudley Knowles, Editor

The Rise of Neo-Kantianism
German Academic Philosophy Between Idealism and Positivism
Klaus Christian Köhnke

The Fixation of Belief and Its Undoing
Changing Beliefs Through Inquiry
Isaac Levi

Substance and Predication in Aristotle
Frank A. Lewis

Matters of Metaphysics
D. H. Mellor

Knowledge and Evidence
Paul K. Moser

Kant: Political Writings
Second Edition
H. S. Reiss, Editor

Philosophical Papers
Volume I: Objectivity, Relativism, and Truth
Volume II: Essays on Heidegger and Others
Richard Rorty

The Post-Modern and the Post-Industrial
A Critical Analysis
Margaret A. Rose

Understanding Action
An Essay on Reasons
Frederic Schick

Automatism, Insanity and the Psychology of Criminal Responsibility
A Philosophical Inquiry
Robert F. Schopp

Morals, Motivation, and Convention
Hume's Influential Doctrines
Francis Snare

Knowledge in Perspective
Selected Essays in Epistemology
Ernest Sosa

Pufendorf: *On the Duty of Man and Citizen According to Natural Law*
James Tully, Editor

Contractarianism and Rational Choice
Essays on David Gauthier's *Morals by Agreement*
Peter Vallentyne, Editor

Political Theory and Postmodernism
Stephen K. White

Hegel: *Elements of the Philosophy of Right*
Allen W. Wood, Editor

The Disappearance of Time
An Essay on the Philosophy of Kurt Gödel
Palle Yourgrau

Available in bookstores or write:

CAMBRIDGE UNIVERSITY PRESS
40 West 20th Street, New York, NY 10011-4211. Call toll-free 800-872-7423.
MasterCard/VISA accepted. Prices subject to change.

TEMPLE UNIVERSITY PRESS

THE METAPHYSICS OF SELF AND WORLD
Towards a Humanistic Philosophy
E. M. ADAMS
"A very impressive, well-written, and philosophically sound work...." —Douglas Browning, University of Texas at Austin
$44.95

THE PASSAGE OF NATURE
DOROTHY EMMET
$39.95

THE RIGHT TO PRIVACY
Gays, Lesbians, and the Constitution
VINCENT J. SAMAR
$39.95

MORAL RESPONSIBILITY AND PERSONS
EUGENE SCHLOSSBERGER
$44.95

THE URGINGS OF CONSCIENCE
A Theory of Punishment
JACOB ADLER
$44.95

STUDIES IN PHILOSOPHY FOR CHILDREN
Harry Stottlemeier's Discovery
Edited by ANN MARGARET SHARP and RONALD F. REED
$24.95/$49.95

ART AND ENGAGEMENT
ARNOLD BERLEANT
$44.95

THE ELEMENTS OF ETHICS
G. E. MOORE
With an introduction by editor Tom Regan
A series of ten previously unpublished lectures by one of the century's most influential philosophers.
$54.95

ANTI-FOUNDATIONALISM OLD AND NEW
Edited by TOM ROCKMORE and BETH J. SINGER
$44.95

EXPLORATIONS AT THE EDGE OF TIME
The Prospects for World Order
RICHARD FALK
"Falk, who has established a long record of being ahead of his time, again pushes our thinking . . . .His ideas are fresh, controversial, provocative." —Robert C. Johansen, Institute for International Peace Studies
$39.95

TWO PATHS TOWARD PEACE
JAMES W. CHILD and DONALD SCHERER
$39.95

JUSTICE, LAW, AND VIOLENCE
Edited by JAMES B. BRADY and NEWTON GARVER
$44.95

IRRATIONALISM
Lukacs and the Marxist View of Reason
TOM ROCKMORE
$44.95

CHRISTIAN PACIFISM
DANIEL A. DOMBROWSKI
"Well-written. . . . A useful statement of an important position." —George Mavrodes, University of Michigan
$34.95

THE CASE AGAINST CHRISTIANITY
MICHAEL MARTIN
$44.95

FAITH, REASON, AND SKEPTICISM
Edited by MARCUS HESTER
Essays by William P. Alston, Robert Audi, Terence Penelhum, and Richard H. Popkin
$39.95

AGAINST INSTINCT
From Biology to Philosophical Psychology
DENNIS M. SENCHUK
$44.95

NOW IN PAPERBACK....

HEIDEGGER AND NAZISM
VICTOR FARIAS
$18.95/$34.95

CAMUS
A Critical Examination
DAVID SPRINTZEN
$18.95/$44.95

THE FORMS OF POWER
From Domination to Transformation
THOMAS E. WARTENBERG
$19.95/$39.95

THE PHILOSOPHY OF ALAIN LOCKE
Harlem Renaissance and Beyond
Edited by LEONARD HARRIS
$17.95/$34.95

THE ROOTS OF THINKING
MAXINE SHEETS-JOHNSTONE
$22.95/$49.95

INTERPRETING WITTGENSTEIN
A Cloud of Philosophy, a Drop of Grammar
RONALD SUTER
$22.95/$44.95

DILEMMAS OF THE AMERICAN SELF
JOHN P. HEWITT
Winner of the 1990 Charles Horton Cooley Award
$18.95/$39.95

TRANSFORMING KNOWLEDGE
ELIZABETH KAMARCK MINNICH
Winner of the 9th Annual Frederic W. Ness Book Award
$18.95/$44.95

THE THEE GENERATION
Reflections on the Coming Revolution
TOM REGAN
$14.95/$24.95

BOOTH 300

order toll-free 1-800-447-1656

TEMPLE UNIVERSITY PRESS

ABORTION RIGHTS AS RELIGIOUS FREEDOM

Peter S. Wenz

$22.95/$49.95

In the *Ethics and Action Series,*
edited by
Tom Regan

"This excellent book is bound to stir debate on the abortion issue and to occupy a rather distinctive position."
—R.G. Frey
Bowling Green State University

"This is an original and scholarly exposition of the view that abortion rights fall under the religion clauses of the First Amendment. . . [and] should help to expand the ethical and constitutional debate about abortion rights."
—Mary Anne Warren, author
Gendercide, The Implications of Sex Selection

BOOTH 300

NO LONGER PATIENT
Feminist Ethics and Health Care

Susan Sherwin

$39.95

"A must for anyone who wants to understand and think clearly about the ethical dimensions of contemporary biomedicine."
—Ruth Hubbard
Professor of Biology, Emerita

"...an important statement."
—Barbara Katz Rothman
Professor of Sociology, Baruch College, CUNY

WOMEN AND GENDER
A Feminist Psychology
Rhoda K. Unger and Mary Crawford
$39.95

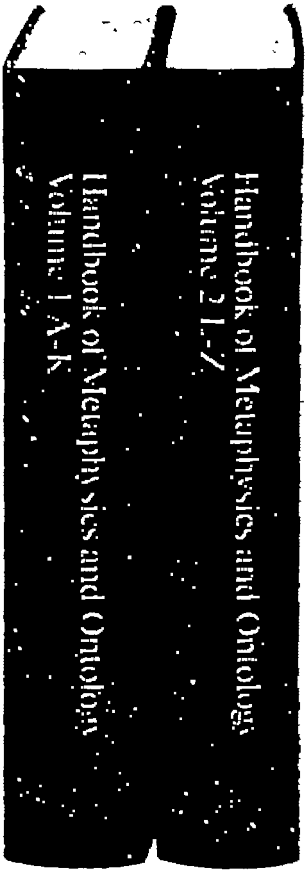
order toll-free 1-800-447-1656

- APA DISCOUNT -

HANDBOOK OF METAPHYSICS AND ONTOLOGY

Editors

Hans Burkhardt Erlangen
Barry Smith Manchester


Handbook of Metaphysics and Ontology
Volume 1 A-K
Handbook of Metaphysics and Ontology
Volume 2 L-Z

Consulting Editors

J. M. Bocheński Fribourg
K. Campbell Sydney
R. M. Chisholm Providence
K. Fine Los Angeles
J. Vuillemin Paris

c. 1.500 pp. hardcover, format:
15.5 x 24.0 cm
publication date: September 1991
ISBN 3-88405-080-X c. $380.00

The work contains more than 450 articles on specific themes and methods in metaphysics and ontology, together with biographical essays summarizing the thinking of the most important philosophers within the tradition.

- classical Greek metaphysics
- medieval and post-medieval Scholasticism
- Renaissance and Enlightenment metaphysics
- rationalism
- the philosophical roots of modern formal logic
- Brentano and his school
- the Polish school
- modern analytic metaphysics
- *plus much more!*

Philosophia

Munich • Philadelphia • Vienna

To Order: P.O. Box 605 • Herndon, VA 22070
(800) 669-7940

APA Booth #416

Prometheus

BOOKS

Prescription: Medicide

The Goodness of Planned Death

Jack Kevorkian, M.D.

Iconoclast Jack Kevorkian, inventor of the controversial "suicide machine," outlines his startling views on planned death and its potential impact on organ harvesting and medical experimentation. Kevorkian made headline in 1990 when he aided Janet Adkins, a victim of rapidly degenerating Alzheimer's disease, in performing the first publicly acknowledged physician-assisted suicide—what he terms *medicide*. Dubbed the "suicide Doctor," he is a man to whom the highest ethical principle is individual self-determination. In this book, he takes on the medical establishment, politicians, and all those who actively resist a rational and comprehensive program of dignified, humane, and beneficial planned death.

245 pages • Cloth $23.95

Animal Experimentation

The Moral Issues

edited by Robert M. Baird and Stuart E. Rosenbaum

This balanced collection focuses attention on the clash of opposing positions in this national debate, whose intensity is sure to increase as demands for more research make the need for rational, informed discussions all the more urgent.

182 pages • Paper $14.95

Metaphysical Delusion

Fraser Cowley

Explains how "linguistic transcendentalism" is employed by philosophers to justify "talking out of both sides of our mouths." Specifically, Cowley demonstrates how empiricism made a metaphysic of science.

145 pages • Cloth $29.95

Introductory Readings in the Philosophy of Science

edited by E. D. Klemke, Robert Hollinger, and A. David Kline

Containing essays by the best theorists in the field, this edition has sections on science and non-science, explanation and law, theory and observation, confirmation and acceptance, science and values, and science and culture.

480 pages • Paper $20.95

Business Ethics

Corporate Values and Society

edited by Milton Snoeyenbos, Robert Almeder, and James Humber

Informative, thought-provoking essays, case examples, and value introductions by the editors combine to form a comprehensive column designed to confront key ethical issues facing contemporary businesspersons and interested readers alike.

502 pages • Paper $21.95

Pornography

Private Right or Public Menace

edited by Robert M. Baird and Stuart E. Rosenbaum

The battlelines are often drawn between those who argue that pornography's damaging social and psychological effects justify censorship, and those who insist that restrictions violate free speech. This collection mirrors these traditional points of view and reflects new directions the debate has taken in recent years, including essays on defining pornography, differences between pornography, erotica, and "moral realism" in artistic expression, the effects of pornography and the growth of a multi-billion dollar industry.

175 pages • Paper $14.95

The Gun Control Debate

You Decide

edited by Lee Nisbet

"*An excellent point of departure for students beginning to study the gun control debate because it avoids the emotional advocacy positions common in other sources.*" —Library Journal

341 pages • Paper $16.95

Clear Thinking

A Practical Introduction

Hy Ruchlis

A lively introduction to the basic principles of critical analysis that shows how to sue sound reasoning skills in everyday situations.

"*It might change your life . . . for the better . . .*" —Isaac Asimov

271 pages • Paper $15.95

Controlling Technology

Contemporary Issues

edited by William B. Thompson

Features classic and contemporary selections that focus on the issues surrounding the nature of technology and the complex ways in which it interacts with human values.

460 pages • Paper $21.95

African-American Humanism

An Anthology

edited by Norm R. Allen, Jr.

This ambitious collection demonstrates the influcence of humanism and free-thought on the history and ideals of black intellectualism.

300 pages • Paper $16.95

Order Toll Free
(800) 421-0351 (24 hours)
or write to
59 John Glenn Drive
Amherst, NY 14228-2197

Add $3.00 p&h for first book, $1.50 for each additional book ($9.00 maximum).
NY State residents add sales tax.

#40122

See these and other titles at booths [illegible]

O X F O R D

Truth and the End of Inquiry
A Peircean Account of Truth
C.J. MISAK
1991 200 pp. $49.95

Reason and Tradition in Indian Thought
An Essay on the Nature of Indian Philosophical Thinking
JITENDRA NATH MOHANTY
February 1992 320 pp. $75.00

Equality and Partiality
THOMAS NAGEL
1991 208 pp. $17.95

What Does It All Mean?*
A Very Short Introduction to Philosophy
THOMAS NAGEL
1987 106 pp. paper $8.95
cloth $16.95

Essays on Aristotle's *De Anima*
Edited by MARTHA C. NUSSBAUM and AMELIE OKSENBERG RORTY
February 1992 464 pp. $69.00

New in paperback!
Love's Knowledge*
Essays on Philosophy and Literature
MARTHA C. NUSSBAUM
Paper February 1992 432 pp.
paper $15.95 cloth $42.50

New in paperback!
Pythagoras Revived*
Mathematics and Philosophy in Late Antiquity
DOMINIC J. O'MEARA
Paper 1991 264 pp. paper $26.00

Hume's System*
An Examination of the First Book of His Treatise
DAVID PEARS
1991 224 pp. paper $14.95
cloth $49.95

Love and Power
The Role of Religion and Morality in American Politics
MICHAEL J. PERRY
1991 240 pp. $29.95

New in paperback!
Reason and Religious Belief*
An Introduction to the Philosophy of Religion
MICHAEL PETERSON, WILLIAM HASKER, BRUCE REICHENBACH, and DAVID BASINGER
1990 304 pp. paper $17.95

New in paperback!
Love and Friendship in Plato and Aristotle*
A.W. PRICE
Paper 1990 280 pp.
paper $19.95 cloth $49.95

New in paperback!
Created from Animals*
The Moral Implications of Darwinism
JAMES RACHELS
Paper 1991 256 pp.
paper $9.95 cloth $22.95

Economics and the Philosophy of Science
DEBORAH A. REDMAN
1990 272 pp. $39.95

New in paperback!
Existential Epistemology*
A Heideggerian Critique of the Cartesian Project
JOHN RICHARDSON
Paper 1991 224 pp.
paper $18.95 cloth $49.95

The Nature of Mind*
Edited by DAVID M. ROSENTHAL
1991 656 pp. paper $19.95
cloth $49.95

OXFORD ◆ UNIVERSITY ◆ PRESS

COME VISIT OUR BOOTHS #306 & #308

Playing by the Rules
A Philosophical Examination of Rule-Based Decision-Making in Law and Life
FREDERICK SCHAUER
1991 272 pp. $59.00

The Mystery of Numbers
ANNEMARIE SCHIMMEL
January 1992 288 pp. $19.95

New in paperback!
The Fabric of Character*
Aristotle's Theory of Virtue
NANCY SHERMAN
Paper 1991 232 pp.
paper $24.95 cloth $45.00

What is Justice?*
Classic and Contemporary Readings
Edited by ROBERT C. SOLOMON and MARK C. MURPHY
1990 394 pp. paper $15.95
cloth $38.00

Autonomy and Rights
The Moral Foundations of Liberalism
HORACIO SPECTOR
1991 216 pp. $49.95

Seven Theories of Human Nature*
Second Edition
LESLIE STEVENSON
Paper 1988 160 pp.
paper $9.95 cloth $24.95

New in paperback!
Freedom and Belief*
GALEN STRAWSON
Paper 1991 360 pp.
paper $22.00 cloth $49.95

New in paperback!
The Existence of God
Revised Edition
RICHARD SWINBURNE
1991 336 pp. $26.00

Revelation*
From Metaphor to Analogy
RICHARD SWINBURNE
February 1992 240 pp.
paper $19.95 cloth $48.00

Quantum Mechanics*
An Empiricist View
BAS C. VAN FRAASSEN
1991 560 pp. paper $29.95
cloth $81.00

The New Art of Autobiography
An Essay on the Life of Giambattista Vico Written by Himself
DONALD PHILLIP VERENE
1991 280 pp. $65.00

The Compact Oxford English Dictionary
in slipcase with reading glass
New Edition
Edited by EDMUND WEINER and JOHN SIMPSON
1991 2416 pp. $295.00

Adam Smith and His Legacy for Modern Capitalism
PATRICIA H. WERHANE
1991 240 pp. $29.95

The Continuous and the Discrete
Ancient Physical Theories from a Contemporary Perspective
MICHAEL J. WHITE
1991 340 pp. $69.00

Demonstratives*
Edited by PALLE YOURGRAU
1990 288 pp. paper $14.95
cloth $49.95

Prices and publication dates are subject to change
Asterisks indicate the book is available for examination. To request an examination copy, write on school letterhead giving full course information, including course name, level, expected enrollment, and your decision deadline, to College Humanities & Social Sciences Marketing Department

200 MADISON AVENUE ◆ NEW YORK, NY 10016

OXFORD

Particles and Waves*
Historical Essays in the Philosophy of Science
PETER ACHINSTEIN
1991 352 pp. paper $24.95
cloth $49.95

Euphemism and Dysphemism
Language Used as Shield and Weapon
KEITH ALLAN and KATE BURRIDGE
1991 288 pp. $24.95

New in paperback!
Eunomia*
New Order for a New World
PHILIP ALLOTT
1991 528 pp. paper $29.95

New in paperback!
The Philosophy of the Visual Arts*
PHILIP A. ALPERSON
1991 528 pp. paper $29.95

Free Will and the Christian Faith
W.S. ANGLIN
1991 232 pp. $55.00

Oxford Studies in Ancient Philosophy*
Volume IX: 1991
JULIA ANNAS
1991 290 pp. paper $29.95
cloth $72.00

In My Father's House
Africa in the Philosophy of Culture
KWAME ANTHONY APPIAH
January 1992 288 pp. $29.95

On Rhetoric*
A Theory of Civil Discourse
ARISTOTLE
Translated by GEORGE A. KENNEDY
1991 368 pp. paper $9.95
cloth $27.95

New in paperback!
Marx's Radical Critique of Capitalist Society*
A Reconstruction and Critical Evaluation
N. SCOTT ARNOLD
Paper 1991 352 pp.
paper $14.95 cloth $35.00

Confessions
SAINT AUGUSTINE
Translated with an Introduction and Notes by HENRY CHADWICK
1991 350 pp. $24.95

Theories of Everything
The Quest for Ultimate Explanation
JOHN D. BARROW
1991 240 pp. $22.95

New in paperback!
The Origins and Development of Classical Hinduism*
A.L. BASHAM
Edited by KENNETH G. ZYSK
1991 208 pp. paper $8.95

Essays on the Philosophy of Socrates*
Edited by HUGH H. BENSON
1991 336 pp. paper $14.95
cloth $45.00

Care and Commitment
Taking the Personal Point of View
JEFFREY BLUSTEIN
1991 288 pp. $32.50

New in paperback!
Plato's Theaetetus*
DAVID BOSTOCK
Paper 1991 296 pp.
paper $24.00 cloth $59.00

The Nature of All Being
A Study of Wittgenstein's Modal Atomism
RAYMOND BRADLEY
January 1992 304 pp. $39.95

OXFORD ◆ UNIVERSITY ◆ PRESS

COME VISIT OUR BOOTHS #306 & #308

International Encyclopedia of Linguistics
4-volume set
Edited by WILLIAM BRIGHT
1991 1856 pp. $395.00 the set

New in paperback!
Pagan Virtue*
An Essay in Ethics
JOHN CASEY
Paper 1991 256 pp.
paper $19.95 cloth $55.00

The Nature of Suffering
And the Goals of Medicine
ERIC J. CASSELL
1991 288 pp. $22.95

New in paperback!
Philosophy, Politics, Autonomy*
Essays in Political Philosophy
CORNELIUS CASTORIADIS
Edited by DAVID AMES CURTIS
1991 320 pp. paper $14.95
cloth $35.00

The Vital Machine
A Study of Technology and Organic Life
DAVID F. CHANNELL
1991 216 pp. $22.95

New in paperback!
Constructibility and Mathematical Existence*
CHARLES S. CHIHARA
Paper 1991 304 pp.
paper $24.95 cloth $49.95

New in paperback!
Philosophical Issues in Journalism*
ELLIOT D. COHEN
January 1992 272 pp.
paper $14.95

Experts in Uncertainty
Expert Opinion and Subjective Probability in Science
ROGER M. COOKE
1991 448 pp. $65.00

The Lopsided Ape
The Evolution of the Generative Mind
MICHAEL C CORBALLIS
1991 384 pp. $24.95

The Kantian Sublime*
The Morality to Art
PAUL CROWTHER
Paper 1991 192 pp.
paper $22.00 cloth $49.95

Theory Change in Science
Strategies from Mendelian Genetics
LINDLEY DARDEN
1991 328 pp. $39.95

New in paperback!
Pragmatics*
A Reader
Edited by STEVEN DAVIS
1991 608 pp. paper $32.50

In Search of Human Nature
The Decline and Revival of Darwinism in American Social Thought
CARL N. DEGLER
1991 416 pp. $24.95

The Critical Theory of Technology*
ANDREW FEENBERG
1991 256 pp. paper $12.95
cloth $32.50

Confrontations with the Reaper
A Philosophical Study of the Nature and Value of Death
FRED FELDMAN
1991 240 pp. $24.95

Philosophical Interpretations
ROBERT J. FOGELIN
1991 240 pp. $24.95

200 MADISON AVENUE ◆ NEW YORK, NY 10016

O X F O R D

New in paperback!
Named an Outstanding Academic Book of 1988-1989 by Choice

Conversations with Nietzsche*
A Life in the Words of His Contemporaries
Edited with an Introduction by
SANDER L. GILMAN
Translated by DAVID J. PARENT
Paper 1991 304 pp. paper $12.95

The Imaginary Museum of Musical Works
LYDIA GOEHR
December 1991 288 pp. $59.00

Drawing the Line
Life, Death, and Ethical Choices in an American Hospital
SAMUEL GOROVITZ
1991 224 pp. $19.95

Law and Objectivity
KENT GREENAWALT
January 1992 304 pp. $34.95

New in paperback!
Religious Convictions and Political Choice*
KENT GREENAWALT
Paper 1991 280 pp.
paper $13.95 cloth $38.00

The Conception of Value
PAUL GRICE
1991 176 pp. $45.00

Russell's Idealist Apprenticeship
NICHOLAS GRIFFIN
1991 424 pp. $98.00

A Daoist Theory of Chinese Thought
A Philosophical Interpretation
CHAD HANSEN
January 1992 496 pp. $57.00

New in paperback!
Founders of Thought*
Plato, Aristotle, Augustine
R.M. HARE, JONATHAN BARNES, and HENRY CHADWICK
1991 320 pp. paper $9.95

New in paperback!
The Language of Morals*
R.M. HARE
1991 206 pp. paper $19.95

New in paperback!
The Concept of Identity*
ELI HIRSCH
Paper February 1992
328 pp. paper $16.50 cloth $34 50

Acceptable Evidence
Science and Values in Risk Management
Edited by DEBORAH G. MAYO and RACHELLE HOLLANDER
1991 288 pp. $35.00

Privacy, Intimacy, and Isolation
JULIE INNESS
March 1992 192 pp. $24.95

Conditionals*
FRANK JACKSON
1991 248 pp. paper $16.95
cloth $49.95

New in paperback!
The Limits of Morality*
SHELLY KAGAN
Paper 1991 432 pp.
paper $24.95 cloth $55.00

New in paperback!
Marx and Ethics*
PHILIP J. KAIN
Paper 1991 232 pp.
paper $19.95 cloth $49.95

OXFORD ◆ UNIVERSITY ◆ PRESS

COME VISIT OUR BOOTHS #306 & #308

Creation and Abortion*
An Essay in Moral and Legal Philosophy
FRANCES MYRNA KAMM
January 1992 192 pp.
paper $9.95 cloth $27.50

Announcing an exciting new reference work!
The Oxford Dictionary of Byzantium
3 volumes
Edited by ALEXANDER P. KAZHDAN
1991 2366 pp. $275.00

Ethics and Civil Rights
RICHARD KING
February 1992 320 pp. $35.00

Character
JOEL J. KUPPERMAN
1991 208 pp. $28.00

Contemporary Political Philosophy*
An Introduction
WILL KYMLICKA
1991 336 pp. paper $13.95
cloth $42.00

New in paperback!
Liberalism, Community, and Culture*
WILL KYMLICKA
Paper 1991 288 pp. paper $19.95

Philosophical Applications of Free Logic
Edited by KAREL LAMBERT
1991 320 pp. $49.95

Morality and Moral Theory*
A Reappraisal and Reaffirmation
ROBERT B. LOUDEN
1991 272 pp. paper $14.95
cloth $35.00

In the Interest of the Governed
A Study in Bentham's Philosophy of Utility and Law
Revised Edition
DAVID LYONS
1991 176 pp. $45.00

Plato on the Self-Predication of Forms
JOHN MALCOLM
1991 240 pp. $55.00

In Defense of the Accidental*
ODO MARQUARD
Translated by ROBERT M. WALLACE
1991 144 pp. paper $12.95
cloth $32.50

The Philosophy of Language
Second Edition
Edited by A.P. MARTINICH
1990 544 pp. $31.00

The Word and the World
India's Contribution to the Study of Language
BIMAL KRISHNA MATILAL
1991 200 pp. $19.95

Springs of Action
Understanding Intentional Behavior
ALFRED R. MELE
1991 280 pp. $39.95

False Consciousness
DENISE MEYERSON
1991 200 pp. $49.95

Forthcoming!
Mind in Context
A Sociocultural Theory of Cognition
Edited by NORRIS MINICK, ELLICE ANN FORMAN, and C ADDISON STONE
May 1992 352 pp. $35.00

200 MADISON AVENUE ◆ NEW YORK, NY 10016

Guilford Publications

Postmodern Theory
Critical Interrogations
Steven Best and Douglas Kellner

"The most detailed introduction to postmodern social theory I have yet seen..."
–Fredric Jameson, Duke University

In this timely volume, the authors analyze postmodern theory to evaluate its relevance for critical social theory and radical politics today. They provide an introduction and critique of the work of Foucault, Deleuze and Guattari, Baudrillard, Laclau and Mouffe, Lyotard, and Jameson; a discussion of postmodern feminist theory and the politics of identity, and a systematic study of the origin of the discourse of the postmodern in historical, sociological, cultural, and philosophical studies.
256 Pages, Paper, ISBN 0-89862-418-5, $16.95, Cloth, ISBN 0-89862-412-6, $36 95

A Theory of Human Need *Len Doyal and Ian Gough*

"This book will have an important place in the academic world It is one of the first of a new crop of books that take up the concerns with need expressed in the 1970s and then abandoned in the 1980s....There is really no book I can think of that tries to combine philosophical analysis and policy in this area in the way this one does."
–Sonia Kruks, Author of Situation and Human Existence

365 Pages, Paper, ISBN 0-89862-419-3, $17.95, Cloth, ISBN 0-89862-413-4, $45 00

Updated and Revised!
Psychoanalytic Politics, 2nd Edition
Jacques Lacan and Freud's French Revolution
Sherry Turkle
327 Pages + Index, Paper, ISBN 0-89862-474-6, $16.95, Cloth, ISBN 0-89862-769-9, $36.95

Available in better bookstores or from

GUILFORD PUBLICATIONS, INC.
72 Spring Street, New York, NY 10012

☎ **CALL TOLL-FREE 1-800-365-7006**
(From Area Codes 212, 718, 516, and 914, call 212-431-9800)

The Journal of Philosophy

Missing Issues?

Complete volumes and all separate issues are available back to Vol. I, No. 1 (January 7, 1904). Prices, not including shipping charges, are as follows:

Single issues:

1904–1976	$3.00
1977–present	$4.00
Dewey Lectures	$5.00

Double issues, devoted to the symposia of the Annual Meetings of the American Philosophical Association, Eastern Division:

1943–1976	$7.00 each set
1977–present	$9.00 each set

Complete volumes:

1904–last year	$60.00 unbound

Indexes:

50-year (1904–1953)	$12.00 cloth
10-year (1954–1963)	$3.50 paper
25-year (1964–1988)	$10.00 paper

Please inquire for shipping charges and discounts on back issues.

709 Philosophy Hall, Columbia University
New York, New York 10027

From philosophy's preeminent textbook publisher...

LOGIC

LOGIC AND CONTEMPORARY RHETORIC: The Use of Reason in Everyday Life, SIXTH EDITION

Howard Kahane

The master is back with hundreds of current new examples and new sections on the pull of pseudoscience, hypothetical question ploy, rhetorical tone, and evaluating irony and humorous rhetoric.

A PRACTICAL STUDY OF ARGUMENT
THIRD EDITION

Trudy Govier

Acclaimed for its integrative, theoretical framework—unique in the field—now rewritten and simplified with more examples and a new Instructor's Manual.

A LOGIC BOOK
SECOND EDITION

Robert M. Johnson

Coverage of traditional logic in a nontechnical format, with two new chapters on inductive logic and extended arguments.

BUILDING ARGUMENTS

Drew E. Hinderer

A simplified, building block approach to reasoning and writing for average students.

INTRODUCTORY SYMBOLIC LOGIC

John K. Wilson

With lively examples and clear, direct language, this text's skill-building presentation offers strategies for approaching formal proofs.

ETHICS

APPLYING ETHICS: A Text with Readings,
FOURTH EDITION

Jeffrey Olen and Vincent Barry

Focuses students on thinking critically with examples, textual discussions and background on issues, point/counterpoint debates, and original texts. New: a theories chapter, a feminist critique of traditional ethics, and a chapter on environmental ethics.

INTERVENTION AND REFLECTION: Basic Issues in Medical Ethics,
FOURTH EDITION

Ronald Munson

Updated throughout including a new chapter on AIDS, coverage of RU 486, "fetus abuse," coverage of the *Webster* and *Cruzan* cases plus new articles by Steinbock, Regan, and Daniels.

MORAL ISSUES IN BUSINESS, FIFTH EDITION

William H. Shaw and Vincent Barry

Thorough updating of the most complete and flexible text in the field, with original text, articles, and over 40 cases.

W Wadsworth

...timeless questions, logical answers

ETHICS

APPLYING MORAL THEORIES
SECOND EDITION

C. E. Harris, Jr.

Helps students analyze moral problems into their component parts and provides a procedure for applying the primary moral theories.

GREAT TRADITIONS IN ETHICS
SEVENTH EDITION

Sheldon Peterfreund and Theodore Denise

Chronological organization and judicious editing of the ethical "greats," with new chapters on Kierkegaard, Sartre, and Annette Baier.

POLITICAL / SOCIAL PHILOSOPHY

PHILOSOPHICAL PROBLEMS IN THE LAW

David M. Adams

Focusing on applied topics and cases, this anthology is a lively and comprehensive introduction to the classic and contemporary issues in legal philosophy.

DEMOCRACY: Theory and Practice

John Arthur

Includes historical and current material on the theory of democracy and its applications to the workplace, capitalism, and the law.

SHOULD WE CONSENT TO BE GOVERNED?
A Short Introduction to Political Philosophy

Stephen Nathanson

Views different responses to political participation and legitimacy from the standpoint of the consumer.

JUSTICE: Alternative Political Perspectives,
SECOND EDITION

James P. Sterba

Covers five different conceptions of justice: libertarian, socialist, welfare liberal, and—new to this edition—communitarian and feminist.

17TH–18TH CENTURY PHILOSOPHY

CENTRAL READINGS IN THE HISTORY OF MODERN PHILOSOPHY:
From Descartes to Kant

Robert Cummins and David Owen

The key writings of the major philosophers of the period, contained in a single useful volume.

For more information
about these or
other philosophy titles,
please write to:

Wadsworth
Publishing Company
10 Davis Drive, Belmont, CA 94002

Wadsworth

From historical to multicultural— five approaches to the introductory course from philosophy's preeminent textbook publisher

ARCHETYPES OF WISDOM: An Introduction to Philosophy

Douglas Soccio

At last, a text that returns to the essence of the philosophical impulse presenting philosophy in the original sense as the search for wisdom. With this powerful interpreting theme, together with a chronological sequence and a storytelling approach, author Doug Soccio offers students the most intriguing introduction to the field they are likely to encounter.

PHILOSOPHY: The Quest for Truth

SECOND EDITION

Louis P. Pojman

Well chosen, representative selections that students find accessible and interesting, along with valuable study aids. Many improvements including a section on applied ethics.

VOICES OF WISDOM: A Multicultural Philosophy Reader

Gary E. Kessler

Unique in its multicultural character, this reader presents classic Western material as well as Hindu, Buddhist, Chinese, African, Latin American, Islamic, Jewish, and feminist perspectives.

OPEN QUESTIONS: An Introduction to Philosophy

Emmett Barcalow

Probing questions, reasoned analysis, contemporary examples, and easy-reading prose incorporating an anthology of classic readings.

INVITATION TO PHILOSOPHY: Issues and Options,

SIXTH EDITION

Stanley M. Honer, Thomas C. Hunt, and Dennis L. Okholm

New co-author Dennis Okholm has revised and updated this successful handbook.

For more information about these or other philosophy titles, please write to:

Wadsworth Publishing Company
10 Davis Drive
Belmont, CA 94002

Wadsworth

2 MAR 1992

Philos P 0

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NUMBER 12
DECEMBER 1991

page
693 *The Role of Trust in Knowledge* John Hardwig
709 *Signal, Decision, Action* Peter Godfrey-Smith

BOOK REVIEWS
723 Colin McGinn: *Mental Content* Pierre Jacob
728 Jon Elster: *Cement of Society* Jean Hampton

739 INDEX TO VOLUME LXXXVIII

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Arthur C. Danto, Kent Greenawalt, Charles Larmore, Isaac Levi, Mary Mothersill. *Special Projects Editor:* Sidney Morgenbesser. *Consulting Editors:* Akeel Bilgrami, James T. Higginbotham, Shaughan Lavine, Charles D. Parsons. *Managing Editor:* Michael Kelly.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President,* Arthur C. Danto: *Vice President,* Corliss Lamont; *Secretary,* Michael J. Mooney; *Treasurer,* Barbara Gimbel; *Other Trustees,* Leigh S. Cauman, Kent Greenawalt, André Schiffrin.

All communications to the editors and trustees and all manuscripts may be sent to Michael Kelly, Managing Editor, 709 Philosophy Hall, Columbia University, New York, NY 10027. FAX: (212) 932-3721.

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL'S terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying More information on our terms may be obtained by consulting our January issue or by writing to us.

© Copyright 1991 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER Second class postage paid at New York, NY, and additional mailing offices Send address changes to *Journal of Philosophy,* 709 Philosophy, Columbia U, NYC 10027

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXVIII, NO. 12, DECEMBER 1991

THE ROLE OF TRUST IN KNOWLEDGE*

The whole fabric of research is trust.

Elizabeth Neufeld

It seems paradoxical that scientific research, in many ways one of the most questioning and skeptical of human activities, should be dependent on personal trust. But the fact is that without trust the research enterprise could not function. . . . Research is a collegial activity that requires its practitioners to trust the integrity of their colleagues.

Arnold S. Relman

We do not normally notice the air we breathe. Similarly, we epistemologists have not noticed the climate of trust that is required—or so I shall argue—to support much of our knowledge. Thus, the title of this paper may seem strange, for most epistemologists and philosophers of science see no role for trust in knowledge. Although epistemologists debate various theories of knowledge, almost all seem united in the supposition that knowledge rests on evidence, not trust. After all, trust, in order to be trust, must be at least partially blind. And how can knowledge be blind? Thus, for most epistemologists, it is not only that trust plays no role in knowing; trusting and knowing are deeply antithetical. We can not know by trusting in the opinions of others; we may have to trust those opinions when we do not know.

I shall argue that this is badly mistaken. Modern knowers cannot be independent and self-reliant, not even in their own fields of specialization. In most disciplines, those who do not trust cannot know;

---

* I wish to thank James O. Bennett, Kathleen Bohstedt, George Brenkert, E. Roger Jones, John Nolt, Dan Turner, and especially Mary Read English and the members of the Philosophy Department at East Tennessee State for helpful comments. Research for this paper has been supported by the National Endowment for the Humanities and by East Tennessee State University.

0022-362X/91/8812/693–708

those who do not trust cannot have the best evidence for their beliefs. In an important sense, then, trust is often epistemologically even more basic than empirical data or logical arguments: the data and the argument are available only through trust. If the metaphor of foundation is still useful, the trustworthiness of members of epistemic communities is the ultimate foundation for much of our knowledge.

I think my argument is applicable to many areas of knowledge. I shall take science and mathematics as my paradigms, however, for they have provided the primary models of knowing for Western epistemology for the last 350 years. I shall attempt to show how and why trust is essential to scientists and mathematicians, and assume that, if I can show this, most epistemologists will agree that we must make room in our epistemologies for trust.

The conclusion that much of our knowledge rests on trust will, I believe, have far-reaching implications. It may force basic changes in epistemology and the philosophy of science. But it is worth emphasizing that I am not here proposing a new epistemology, nor endorsing a "nonstandard" analysis, such as that of Lorraine Code,[1] Michael Welbourne,[2] or the "strong programme" of the Science Studies Unit at Edinburgh. Quite the opposite: I wish to address my argument to as many philosophers as possible. I therefore aim to work within the standard analyses of knowledge and of science. My purpose here is to call attention to a feature of modern science and mathematics which has not received sufficient attention. I leave for another occasion the large question of whether accepting my conclusion would force basic changes in epistemology and the philosophy of science or whether the idea of knowledge based on trust could be assimilated by the received views.

I

In the early 1960s, Derek de Solla Price[3] observed that there was a rapid trend away from single-author papers in scientific journals. In fact, the trend is toward an ever-increasing number of authors per article.[4] Modern science is collegial not only in the sense that scientists build on the work of those who have preceded them, but also in the sense that research is increasingly done by teams and, indeed, by larger and larger teams. This is true for two reasons.

---

[1] *Epistemic Responsibility* (Hanover: New England UP, 1987).

[2] *The Community of Knowledge* (Aberdeen, Scotland: University Press, 1986).

[3] *Little Science, Big Science* (New York: Columbia, 1963), pp. 87–8.

[4] Limitations of space prevent giving references for statements I make about the current state of science or about the opinions of the scientific community. Readers interested in my sources should write me for a documented version of this paper.

(1) The process of gathering and analyzing data sometimes just takes too long to be accomplished by one person. In an earlier paper, I discussed an experiment that measured the lifespan of charm particles.[5] The paper reporting the results of this experiment has 99 authors, in part because it took about 280 person/years to do the experiment. Moreover, even for experiments that require less than a lifetime to run, the pace of science is often far too rapid for a lone experimenter to make any contribution at all by doing them. One of the authors of this paper on charm particles, William Bugg, estimated at the time of publication that, within three years, some other team would come up with a technique that gives a considerably better measurement and that, within five years, the paper would no longer be of interest. Five years later, his prediction has been borne out.

(2) Even more important for the purposes of epistemological analysis, research is increasingly done by teams because no one knows enough to be able to do the experiment by herself. Increasingly, no one could know enough—sheer limitations of intellect prohibit it. The cooperation of researchers from different specializations and the resulting division of cognitive labor are, consequently, often unavoidable if an experiment is to be done at all. No one particle physicist knows enough to measure the lifespan of charm particles. Indeed, Bugg reports that no one university or national laboratory could have done their experiment. None of the authors of such a paper is in a position to vouch for the entire content of the paper.

Teamwork is pretty standard fare within the empirical sciences. But it is not completely unknown in mathematics as well, due to the many areas of specialization required to complete some proofs. As just one example, consider J. Korevaar's[6] recounting of a critical stage in Louis de Branges's proof of Ludwig Bieberbach's conjecture (a conjecture in complex analysis which dates back to 1916, but which had resisted proof for almost 70 years):

> For relatively small n, de Branges could verify immediately that the sums . . . are positive on $(0, \infty)$. But what about larger values of n? At this stage de Branges went to his numerical colleague Gautschi at Purdue University for help. He told Gautschi that he had a way of proving

---

[5] "Epistemic Dependence," this JOURNAL, LXXXII, 7 (1985): 335–49.

[6] "Ludwig Bieberbach's Conjecture and Its Proof by Louis de Branges," *The American Mathematical Monthly*, XCIII (1986): 505–14. I owe this reference to Carl Wagner, who also informs me that there is now a much simpler proof of the Bieberbach conjecture which avoids reliance upon Askey's work.

> the Bieberbach conjecture, but needed to establish certain inequalities involving hypergeometric functions. Would Gautschi be willing to check as many of these inequalities as possible on the computer? Gautschi wrote a suitable program with a feeling that he might soon hit a value of *n* for which the consistent positivity of expressions . . . would come to an end. Much to his surprise, however, he discovered that the crucial expressions were positive for all values of *n* which he tried: $2 \leq n \leq 30$. Thus at this time, assuming that the theoretical work was correct, de Branges and the computer had verified the Bieberbach conjecture for all *n* up to 30!
>
> How to continue? Gautschi had the idea to call Askey at the University of Wisconsin, the world's expert on special functions. At first Askey was incredulous that the supposed positivity of sums . . . would prove the Bieberbach conjecture. However, he realized very soon that those sums were essentially generalized hypergeometric functions of a very special kind which are known to be positive (*ibid.*, pp. 512–3).

Specialization and teamwork are thus inescapable features of much modern knowledge acquisition. This point is not merely a genetic point about "the context of discovery." Classical epistemology can admit—though usually not much is made of this fact—that trust plays a role in the origins of someone's knowledge. But specialization and teamwork apply in the "context of justification" as well. It is quite likely that no one mathematician has or will ever have the logical justification for each step in de Branges's proof. Those (like Askey) who know enough about hypergeometric functions probably do not know enough complex analysis to verify other parts of the proof; those (like de Branges) who know enough complex analysis have not mastered Askey's work. Possibly one mathematician could learn enough different mathematical specialties to grasp each step in de Branges's proof. But mathematicians do not think that would be a particularly useful thing to do, especially since no one can learn everything in mathematics. (Askey's work on special functions is apparently a small niche.) And clearly one particle physicist could never have sufficient justification for any claim at all about the lifespan of charm particles.

What are we epistemologists to say about de Branges's proof or the 99 physicists' measurement of the lifespan of charm particles? I think cases like these force us to make very basic choices: either we can modify our epistemological theories, or we can cling to them and deny that they could possibly be cases of knowing, since they fail to meet our requirements.

The latter option would be to say that experiments or "proofs" requiring teamwork could not possibly yield knowledge or even rational belief. Why not? Because knowing requires good reasons for believing, and none of the mathematicians or physicists has sufficient reasons (except in testimonial form—more on that later) to accept the conclusions of their papers.

Moreover, it will not do epistemologically to have many tiny shreds of the empirical evidence or fragments of the mathematical proof in many separate minds. For it is the interconnection, the structure of these bits of evidence into a unified whole, that enables them to add up to a justified conclusion about charm particles or the Bieberbach conjecture. Since no one has sufficient evidence to justify the conclusion, there is no one who knows. But there is no knowledge without a knower, so the lifespan of charm particles is not known and could not be known. At least, not by humans. And de Branges has produced nothing that could count as a proof.

If this conclusion is unpalatable—as I think it is—we need an epistemological analysis of research teams, for knowledge of many things is possible only through teamwork. Knowing, then, is often not a privileged psychological state. If it is a privileged state at all, it is a privileged social state. So, we need an epistemological analysis of the social structure that makes the members of some teams knowers while the members of others are not. An analysis of testimony and testimonial evidence will provide a start on this project and also the next step in the present argument about the role of trust in epistemology.

II

It is the testimony of one scientist or mathematician to another that connects the bits of evidence gathered by different researchers into a unified whole that can justify a conclusion. By accepting each others' testimony, individual researchers are united into a team that may have what no individual member of the team has: sufficient evidence to justify their mutual conclusion.

Elsewhere, I have developed an epistemological principle, "the principle of testimony" (*op. cit.*):[7]

> (T) If A has good reasons to believe that B has good reasons to believe p, then A has good reasons to believe p.

This principle is general enough to capture the epistemic structure of appeals by a layperson to the intellectual authority of experts. But we are speaking here, not of laypersons, but of research scientists, and of research scientists within the domains of their own specialties. So if such researchers are sometimes knowers, we must consider a stronger version of the principle of testimony:

---

[7] Cf. also "Evidence, Testimony, and the Problem of Individualism—A Response to Schmitt," *Social Epistemology*, II (1988): 309–21. I owe the name, 'principle of testimony', to Frederick Schmitt.

(T′) If *A* knows that *B* knows *p*, then *A* knows *p*.

The problems epistemologists have seen in testimonial evidence are evident in both T and T′. Our epistemological training leads us to ask: How can *A* know that *B* knows *p* unless *A* herself first knows *p*? How can *A* even know that *B* has good reasons unless she herself has those reasons and knows that they are good reasons? Of course, *A* can learn from *B*, but how can *A* know through *B*?

These are good questions, important questions. But the epistemological requirement implicit in the rhetorical version of these questions would render testimony epistemically useless. In order to ground her knowledge that *p*, *A* appeals to *B*. Why? Clearly, the whole point of appealing to the testimony of others is that they know things we do not. If this were not the case, basing belief on testimony would be pointless at best, hence nonrational or irrational. The appeal to *B* must be able to strengthen *A*'s reasons for believing *p*; *A* now knows *p* at least partly because she knows that *B* knows *p*.

To count as good testimonial evidence for *p*, testimony must be working well (more on that below). But when testimony is working well, belief based on testimony is often not, as traditional epistemology would have it, a poor, second-best substitute for direct evidence. On the contrary, belief based on testimony is often epistemically superior to belief based entirely on direct, nontestimonial evidence. For *B*'s reasons for believing *p* will often be epistemically better than any *A* would/could come up with on her own. If the best reasons for believing *p* are sometimes primarily testimonial reasons, if knowing requires having the best reasons for believing, and if *p* can be known, then knowledge will also sometimes rest on testimony.

Nor is this conclusion always dependent on *A*'s limited competence in the domain of whether or not *p*: the only respectable beliefs anyone has or could ever have about charm particles must be based largely on the testimony of others. In fact, a belief based partly on second-hand evidence will be epistemically superior to any belief based completely on direct empirical evidence whenever the relevant evidence becomes too extensive or too complex for any one person to gather it all. For in all such cases, one can have sufficient evidence only through testimony. We are thus driven to accept the stronger principle of testimony: if *A* knows that *B* knows *p*, then *A* knows *p*. We must modify our epistemologies to make them compatible with this principle.

Testimonial evidence has potential problems as well as strengths, however, and they arise from the same feature: in order for testimony to be useful, *A* cannot already have *B*'s reasons. So, if *A* accepts *p* on *B*'s say-so, those reasons (*B*'s reasons) which are necessary to justify *A*'s belief are reasons which *A* does not have. Sometimes it is feasible for *B* to share with *A* all the evidence necessary to justify the claim that *p*. But usually not. Indeed, if *A* and *B* come from different disciplines or even different specialties within the same discipline, *A* often will not know what *B*'s reasons are, much less why they are good reasons for believing *p*.

Thus, the blindness of *A*'s knowledge that *p*: those reasons which are necessary to justify *p* (and *A*'s belief that *p*) are reasons which *A* does not have. Obviously, since she lacks part of the evidence that justifies the claim that *p*, *A* is limited in the extent to which she can effectively scrutinize or challenge *B*'s claim about *p*. And yet we are to say that *A* knows that *p*, despite this blindness, this lack of the evidence necessary to justify *p*, this inability to evaluate the case for *p*?

Strange as this may seem, this is what we must say, unless we wish to maintain (1) that there can no longer be knowledge in many scientific disciplines because there is now too much available evidence (!); (2) that one can know *p* only by ignoring most of the best evidence for *p* (!); or (3) that some knowledge is known by teams or communities, but not by any individual person. Although I believe this third option has more plausibility than is generally acknowledged (it may also be the view of C. S. Peirce and John Dewey), in this paper I shall continue to pursue the idea that *A* does know *p* and that we need to modify our accounts of knowledge and rational belief to account for *A*'s knowledge.

Now, given the fact that *A* does not/cannot have *B*'s reasons, *A*'s position is really this:

(1) *A* knows that *B* says that *p*.
(2) *A* believes (and has good reasons to believe?[8]) that *B* is speaking truthfully, i.e., that *B* is saying what she believes.
(3) *A* believes (and has good reasons to believe?) that *B* (unlike *A*) is in a position, first, to know what would be good reasons to believe *p* and, second, to have the needed reasons.
(4) *A* believes (and has good reasons to believe?) that *B* actually has good reasons for believing *p* when she thinks she does.

---

[8] The reason for the parenthetical question in this statement and the next two is that I am undecided about what to say about implicit trust. If *A* trusts *B* implicitly, she will often not have or even feel the need to have good reasons to believe what *B* says. I think epistemic communities in which the climate of implicit trust prevails have real advantages over those in which good reasons for trusting are felt to be needed and are then supplied.

Although obvious, it is important to note two things about *B* and her contribution to *A*'s good reasons. First, unless *B* believes what she is saying, *B*'s knowledgeability about *p* will not give *A* good reasons to believe *p*. Thus, *A*'s good reasons depend on whether *B* is truthful, or at least being honest in this situation.

Second, even *B*'s truthfulness will not give *A* good reasons to believe *p* if *B* believes she has good reasons when she does not. So, in addition to being truthful, *B* must, first, be competent—she must be knowledgeable about what constitutes good reasons in the domain of her expertise, and she must have kept herself up to date with those reasons. Second, *B* must be conscientious—she must have done her own work carefully and thoroughly. And third, B must have "adequate epistemic self-assessment"—*B* must not have a tendency to deceive herself about the extent of her knowledge, its reliability, or its applicability to whether *p*.

Although the usefulness and the rationality of belief based on testimony stem from the fact that *A* does not, often even cannot, have *B*'s reasons for believing *p*, this fact also reveals that *A*'s reliance on *B*'s testimony must include reliance on *B*. The reliability of *A*'s belief depends on the reliability of *B*'s character. *B*'s truthfulness is part of her moral character. Competence, conscientious work, and epistemic self-assessment are aspects of *B*'s "epistemic character." I shall return to this point frequently, and 'character' will refer to these moral and epistemic qualities. (Although competence is not a character trait per se, it standardly depends on character: becoming knowledgeable and then remaining current almost always requires habits of self-discipline, focus, and persistence.)

In short, *A* must TRUST *B*, or *A* will not believe that *B*'s testimony gives her good reasons to believe *p*. And *B* must be TRUSTWORTHY or *B*'s testimony will not in fact give *A* good reasons to believe *p*, regardless of what she might believe about *B*. A team of scientific experimentalists, for example, must both trust each other and be worthy of that trust or their experiment will not give anyone enough good reasons to believe their conclusions.

We thus reach another epistemologically odd conclusion: the rationality of many of our beliefs depends not only on our own character but on the character of others as well; the rationality of many of our beliefs depends on what others do and hence is not within our individual control. This is perhaps not a strange conclusion when we think about the various ways in which we as laypersons depend on others to have the evidence that supports our beliefs. But it becomes much stranger when we realize that this dependence on the character of others applies even to some of the epistemically best beliefs, i.e., to those of the top experts within their own fields of expertise.

The oddness stems, I contend, from the individualistic bias of most epistemology—with its penchant for epistemic self-reliance and self-sufficiency, and its flight from any form of epistemic vulnerability. But if our epistemic authorities are unreliable, we simply have no alternative but to hold less rational beliefs. Either we must then accept the testimony of unreliable authorities or we must rely on our own relatively inexpert and uninformed judgments.

Now, if *B* must be reliable in order for her testimony to be reliable, it seems that *A* must know *B*—at least to the extent of knowing that *B* is both morally and epistemically reliable—before *A* can have good reasons for believing *p* on the say-so of *B*. But scientists usually must rely on scientific testifiers who are not personally known to them. Clearly, this is often true for scientific testimony embodied in the literature. It is sometimes true even among members of a research team. (The team that measured charm particles was scattered over three continents, and Bugg reports that he knows only 10 or 12 of his 98 coauthors well enough to be able to form any judgment about the quality of their work.)

If *B* is not personally known to *A*, there are two strategies for attempting to ascertain the reliability of *B* and thus her testimony. The first is to check with someone who does know *B* and the quality of her work. This strategy can be expressed by extending the principle of testimony:

> A has good reasons for believing *C* (also *D*, *E*, . . .) has good reasons for believing *B* has good reasons for believing *p*.

Normally, however, *A* will not be in a position to ascertain *C*'s reliability as a testifier about *B*, and *C* will often rely on the testimony of still others in order to form her judgment about *B*. This is not merely a philosophical quibble: as we shall see, the difficulty of gathering dependable testimony about the reliability of *B* is a problem of considerable practical import in contemporary science. Still, by repeating this procedure, i.e., checking with several knowledgeable people in *B*'s field, *A* will be able to ascertain *B*'s reputation within *B*'s discipline, and this surely will give *A* some evidence about the reliability of *B* and her testimony. But this process does not obviate the need for trust—it only redistributes and refines that trust.

There is a second strategy for attempting to ascertain the reliability of *B* when *B* is not personally known to *A*. That is to get a second opinion about the truth of what *B* has said. Often, *A* can find a *C* (also *D*, *E*, . . .) who is independent of *B* and who is also knowledgeable about whether *p*. If *C*, *D*, and *E* corroborate *B*'s testimony, *A* will have better reason to believe it.

We shall return to these strategies below, for they have played a central role in explanations of the special reliability of scientific testimony and, consequently, of the scientific process itself. The second, under the rubric of "replication of experimental results," has been especially important.

III

I have claimed that trust in the testimony of others is necessary to ground much of our knowledge, and that this trust involves trust in the character of the testifiers. But there is a basic objection to this thesis: that prudential considerations alone are sufficient to guarantee that the members of a scientific community will be truthful and also constantly vigilant against self-deception.

Michael Blais has developed this objection in his paper, "Epistemic Tit for Tat."[9] Blais argues that trustworthiness can be modeled as a strategy—indeed, the only prudent strategy—for a member of a scientific community. Blais acknowledges that trust is essential to science, because cooperation is essential, but he maintains that the type of cooperation at work does not require trust in the moral sense. "Only cooperation, as defined . . . in game theory and as illustrated in the Prisoner's Dilemma, is necessary for the justification of vicarious knowledge" (*ibid.*, p. 370). In science, cooperation means not defecting in the knowledge game—in other words, not cheating by fudging, fabricating, or otherwise publishing unreliable results. "Defection means succumbing to the temptation of leaving the other players in the knowledge game with the sucker's pay-off, while attempting to maximize immediate gain" (*ibid.*, pp. 370–1).

Blais echoes the common faith of scientists and philosophers that peer review and replication of results will detect defectors. "What count are factual results that are reproducible. If results cannot be reproduced, they may simply be rejected" (*ibid.*, p. 371). "Peer review and blind refereeing ensure that, in the long run, defectors should be found out" (*ibid.*, p. 372). Since Blais maintains that, in the game of science, the punishment for defecting is permanent exclusion from the game, he concludes that defecting is very imprudent.

At least one scientific community, however, the biomedical research community, has had its faith in replication and peer review shattered by a number of spectacular and highly publicized examples of research fraud.[10] Within biomedical science, the names of the fraudulent researchers have become well-known—John Darsee, Robert Slutsky, John Long, Vijay Soman, William Summerlin, Mark Spector, Stephen Breuning, and now Thereza Imanishi-Kari. Although there are other sloppy, careless, or deceptive research practices that may be even more damaging to the reliability of scientific testimony, "scientific misconduct"—commonly defined as plagiarism or the fabrication, falsification, or deliberate misrepresentation of data—is the most blatant example of defection in the knowledge game. The phenomenon of scientific misconduct reveals that a more thorough-going trust than mere strategic trust is involved in science. The consensus within the biomedical sciences is that neither peer review nor replication is likely to detect careless, sloppy, or even fraudulent research.

---

[9] This JOURNAL, LXXXIV, 7 (1987): 363–75. Blais has reaffirmed this view in a more recent article, "Misunderstanding Epistemic Tit for Tat: Reply to John Woods," this JOURNAL, LXXXVII, 7 (1990): 369–74.

[10] No one knows how widespread fraud in science is. But a survey mailed to more than 2,100 scientists in six fields—physics, chemistry, biology, economics,

The number of really well-qualified referees for peer reviews is often inadequate, given the quantity of articles submitted and the complexity and multiplicity of techniques involved in research. Furthermore, an internally consistent and plausible fabrication cannot be detected by referees, since they do not examine the original data or the gathering of that data. Slutsky's fraud, for example, was detected only because of his statistical naivete and also his very bad luck (two of his papers were read in quick succession by an astute reader). Those who investigated the Slutsky case maintain that his fraudulent papers could have been read independently for years without arousing any suspicion.

Nor will careless or fraudulent research normally be detected by replication, for the structure of modern science acts to prevent replication, not to ensure it. It is virtually impossible to obtain funding for attempts to replicate the work of others, and academic credit normally is given only for new findings.

When replication is attempted, it will not always detect fraudulent papers. In fact, replication paradoxically will support rather than unmask those fraudulent papers which happen to have correct conclusions. Thus, fraudulent papers announcing results that are predictable extensions of basic work done by others are quite likely to pass the test of replication. Even when attempted replication fails to produce similar results, there are often other explanations. Thus, when a group of Swiss researchers failed to replicate some of Dar-

---

psychology, and sociology—revealed that nearly one-quarter of the respondents personally knew someone who had falsified data and 2/5 believed that some of their own work had been plagiarized (M. Davis, *The Perceived Seriousness and Incidence of Ethical Misconduct in Academic Science*. Unpublished Ph.D. thesis. The Ohio State University, 1989).

see's work, they considered various explanations for their inability to confirm his findings, but they did not consider the possibility of fraud. And both Darsee and Spector argued that their experiments were too delicate to work unless the experimenter had extremely subtle skills. For these reasons, Relman,[11] a respected observer, has concluded that "fraudulent data may be rapidly identified in an area of great importance where research activity is intense, but that is probably not true in most fields" (*ibid.*, p. 1416).

We can see why Blais's argument must fail. Game-theoretic arguments such as Blais's rest on two assumptions. The first, which Blais acknowledges, is that the relationship be durable: "a cooperative strategy . . . has little chance in the short or medium term; only long-term relationships permit it to hold its own" (*op. cit.*, p. 368). The second assumption is, obviously, that the other players will recognize when they have received the sucker's pay-off.

These assumptions are the Achilles' heel of Blais's application of game theory to the issues of scientific testimony. Usually, the cooperative relationship is precisely the short- or medium-term relationship. By the standards of game theory, 20 or 30 trials is not a very long run, but a researcher can substantially enhance her career if she can successfully publish several fraudulent articles. Also, as we have seen, it is often very difficult for others to detect defectors. Indeed, there is a class of lapses from acceptable scientific practice which are not discoverable by anyone else, since only the researchers have the actual empirical data upon which their paper is, presumably, based. Even if lab logbooks are kept in a form that allows inspection by others, one would need to have been present during the experiment to know that the logbooks are not themselves fraudulent.

Reliance upon inside informants—co-investigators or others in the lab—is widely recognized to be crucial to the detection of scientific misconduct. But this reliance is also problematic, for many reasons. Detection is often very difficult for those not in the defector's field of specialization, which compounds the problem in multidisciplinary work. Moreover, co-investigators are often not the ones who will be damaged by the defection, since the prevailing ethos within science is that joint authors are not responsible and should not be penalized for the fraudulent or sloppy work of one member of their team. (Consequently, there is a temptation to collaborate with someone you personally know to be a defector—to leave, for example, the collection of data to someone you know to be "sloppy.") Finally, there is a whole range of deterrents to informing on one's colleagues in the lab, ranging from loyalty, to reluctance to meddle in others' affairs, to the absence of adequate protection for scientific whistle blowers. Virtually all observers agree that the confidentiality of an informant cannot be successfully protected. Some informants have paid dearly for their work in uncovering scientific fraud.

---

[11] "Lessons from the Darsee Affair," *New England Journal of Medicine*, CCCVIII, 23 (1983): 1415–7.

Thus, contrary to what Blais suggests, detection is often quite difficult. And, contrary to what Blais maintains, the punishment for proven defection is also often not severe. Although fraudulent researchers whose cases attract wide publicity may well forfeit their reputations and careers, others do not. Some quietly relocate to other institutions; others do not even lose their jobs. A recent survey suggests that scientists themselves do not believe that even temporary exclusion from research would result from proven fraud.[12]

There are reasons, of course, for less than Draconian punishment of research fraud. These include fear of liability, of unfairly tainting the reputations of co-authors, of shaking public confidence in science, or of jeopardizing the reputation and the funding of the laboratory and university in which a fraudulent researcher worked. Since a fraudulent researcher's institution can achieve the penalty of permanent exclusion from the scientific community only at the cost of a great deal of effort and unfavorable publicity, institutions have a strong incentive to settle for less severe and more private punishments.

Clearly, a researcher's prudential concerns are often insufficient to ensure her trustworthiness and thus the reliability of her testimony. General recognition of this fact naturally leads to two responses: (1) deterrents to defection should be strengthened, even if they may be insufficient; (2) if prudential reasons are not sufficient to assure trustworthy testifiers, ethical researchers are required. The need for "research ethics" is now widely acknowledged. The Institute of Medicine—sister to the National Academy of Sciences—has, for example, recently called for universities to provide formal instruction in research ethics for all science students at both the undergraduate and graduate level.

Scientists working in some other fields maintain that attempts to replicate experimental findings are still common in their sciences and thus that "it can't happen here." Perhaps there really is more replication in some sciences than in others, but it seems reasonable to conjecture that there will be less replication (1) of long, costly, or very time-consuming experiments; (2) of experiments that require exotic equipment, substances, or specimens; (3) in those fields in which senior scientists can train their graduate students (and postdocs, residents, fellows) by having them gather new data rather than replicate existing results; (4) of experiments requiring the cooperation of multidisciplinary teams. It also seems likely that (5) considerations of the ethical use of research subjects will ensure that there will be no routine or even widespread replication of experiments involving human—or, increasingly, even animal—research subjects. These conditions apply to many sciences—and they apply increasingly as scientific research becomes more sophisticated, complex, and costly.

---

[12] J. P. Tangney, "Fraud Will Out—Or Will It?" *New Scientist*, cxv (August 6, 1987), pp 62–3.

Peer review and attempted replication may once have been effective deterrents to fraudulent or deceptive publication. They are now much weaker deterrents. In any case, they can never have been completely effective detectors of fraudulent or misleading testimony, for, as we have seen, both have in-principle weaknesses.

IV

Often, then, a scientific community has no alternative to trust, including trust in the character of its members. The modern pursuit of scientific knowledge is increasingly and unavoidably a very cooperative enterprise. Cooperation, not intellectual self-reliance, is the key virtue in any scientific community. But epistemic cooperation is possible only on the basis of reliance on the testimony of others. Scientific propositions often must be accepted on the basis of evidence that only others have. Consequently, much scientific knowledge rests on the moral and epistemic character of scientists. Unavoidably so. Not because "hard data" and logical arguments are not necessary, but because the relevant data and arguments are too extensive and too difficult to be had by any means other than testimony.

Indeed, a scientific community is often forced to rely on the testimony of one testifier. Replication of experiments is not standard. Moreover, peer review can never detect plausible and internally consistent fabrications, and attempted replication will support rather than unmask fraudulent reports that happen to have true conclusions. So, there sometimes simply are no *C*, *D*, and *E* who can corroborate or effectively challenge the testimony of *B*. And the testimony of *C*, *D*, and *E* about the professional reputation of *B* may well be seriously misleading: before their frauds were detected, some fraudulent biomedical researchers were widely regarded within their disciplines as outstanding exemplars of scientific productivity, even as scientific geniuses.

Institutional reforms of science may diminish but cannot obviate the need for reliance upon the character of testifiers. There are no "people-proof" institutions. And even if it were possible fully to police scientific research, it would still be necessary to rely on the integrity of the newly-created "science cops" and the reliability of their testimony. (There has, for example, been Congressional concern over the lack of independence of those now being assigned the responsibility of investigating research fraud.)

Science, then, is not completely different from other cooperative enterprises: the reliability of scientific testimony, like the reliability of most other testimony, ultimately depends on the reliability of the testifier, or on the reliability of those charged with ensuring the reliability of the testifier. Obviously, none of this means that deterrents to scientific misconduct are unimportant. Since knowledge of right and wrong will not deter the unscrupulous, deterrents remain important, even if insufficient. Still, the main point here is that scientific knowledge rests on trust, and not only on trust in deterrents, but also on trust in the character of scientists.

The view of Blais (and many, many others) thus becomes part of the problem: one reason the problem of fraudulent research looms so large is that for decades scientists have insisted that science is virtually fraud-proof. Philosophers have too often joined in the chorus. Inability to see the role of trust in science effectively destroys our ability to combat unreliable scientific testimony. It undermines any attempt to formulate and teach research ethics and it stifles any attempt to introduce new deterrents to fraud. A fraud-proof institution has no need for additional protection against fraud.

We must, however, guard against jumping to the conclusion that we should try to redesign our epistemic institutions so as to minimize the role of trust in knowledge. For the alternative to trust is, often, ignorance. An untrusting, suspicious attitude would impede the growth of knowledge, perhaps without even substantially reducing the risk of unreliable testimony. Trust in one's epistemic colleagues is not, then, a necessary evil. It is a positive value for any community of finite minds, provided only that this trust is not too often abused. For finite minds can know many things only through epistemic cooperation. There is, then, a very delicate balance between places for trust within epistemic communities and places for insisting on better safeguards against untrustworthy testifiers.

Philosophers of science, epistemologists, logicians, and ethicists all should be helping concerned members of scientific communities both to formulate an ethics of scientific research and testimony, and also to design structural reforms that will strengthen the deterrents to scientific misconduct without destroying collegiality and creativity. But we may not be able to help until we get our own conceptual houses in better order. The conclusion that knowledge often is based on certain kinds of relationships between people, on trust, and consequently on the character of other people is an epistemologically odd conclusion. It is odd even for pragmatists, though pragmatists have generally had more room for epistemic community in their theories. To my mind, this oddness is symptomatic of what will be needed to assimilate an acknowledgement of the role of trust into our epistemologies. We have a lot of work to do.

Clearly, the implications of the role of trust in knowledge will reach beyond epistemology and the philosophy of science into ethics and social philosophy. I close with just one example. The prevailing tenor of twentieth-century Anglo-American philosophy has been that epistemology is more basic than ethics. On this view, ethics must meet epistemological standards on pains of bankruptcy. And the prevailing suspicion in our culture—a suspicion nurtured by philosophy—is that ethics cannot pass the epistemological test, and that there is thus no ethical knowledge. Science, in contrast, is commonly believed to be too "hard" and "objective" to require anything as mushy and subjective as ethics.

But scientific realism—indeed any theory that grants objectivity to scientific judgments—turns out to be incoherent when combined with subjectivism or skepticism in ethics. It remains true, of course, that ethical claims must meet epistemological standards. But if much of our knowledge rests on trust in the moral character of testifiers, then knowledge depends on morality and epistemology also requires ethics. In order to qualify as knowledge (or even as rational belief), many epistemic claims must meet ethical standards. If they cannot pass the ethical muster, they fail epistemologically.

JOHN HARDWIG

East Tennessee State University

SIGNAL, DECISION, ACTION*

Consider an inner state guiding a particular behavior in an idealized animal, and an external condition to which this behavior is an appropriate response. The inner state is also linked to the idealized animal's senses. What kinds of statistical relations will obtain among the action-steering inner state, the animal's senses, and the world when the animal is wired up in an optimal way?

The behavior I am discussing is best performed when, and only when, the important external condition obtains. So in a sense, the optimal inner state is one that is tokened when and only when the external state obtains. But I shall assume that this kind of optimality is impossible for the animal, because it does not live in an epistemically benign world. Although the important environmental state does have characteristic effects on the animal's senses, it is possible for random "noise" in the environment to produce some of these sensory states as well. So, if the inner state is thought of as a belief about the state of the world, it is not possible for the animal always to believe truly. The topic of this paper is the animal's optimal wiring, where this optimality is understood as relative to a set of constraints imposed by imperfect senses and an unfriendly environment.[1]

I suggested that the inner state could be thought of as a belief about the world. This idea will have heuristic value at a number of points in the discussion, but it is not essential to the model. Some might prefer to think of the inner state as a decision to act, others might refuse to interpret it in this kind of way at all. As one of the products of this discussion will be an outline of a naturalistic semantic theory, it is important that everything in the model can be described nonintentionally. But there are other ways to put the model to philosophical work, and many of these projects are more easily

---

* This paper has benefitted from audiences at the University of Pennsylvania, UC Berkeley, the 1990 conference of the Australasian Philosophical Association, the University of Maryland, and Rijeka and Zadar (Yugoslavia). I am especially grateful, however, to Philip Kitcher. I would also like to thank the University of Sydney for financial support.

[1] For discussions of other aspects of the adaptive wiring of simple model animals, see Valentino Braitenberg, *Vehicles: Experiments in Synthetic Psychology* (Cambridge: MIT, 1984); and Paul Churchland, "Cognitive Neurobiology: A Computational Hypothesis for Laminar Cortex," *Biology and Philosophy*, I, 1 (1985): 25–51.

0022-362X/91/8812/709–22

presented if the question about reducing semantic properties is ignored much of the time.

There are five schematic letters that must be introduced at this point. Four of them refer to state types, and one to a variable. They are most easily remembered with the aid of a diagram (figure 1).

The important environmental condition is S (for "signal"). The other possible state of the world is N (for "noise"). X is a variable describing the animal's sensory input. C (for "cause") is the action-guiding inner state of the animal, and M is the movement or behavior which C produces. In this discussion, I shall assume that the link between C and M is fixed: C always causes M, and nothing else does. I shall also assume certain fixed relations between input X and the two-world states S and N; these will be described shortly. The connection that can be tuned is the link between X and C; the question is about which sensory events should cause C in an optimally wired animal.

The first outlines of the answer are obvious. The animal wants to produce M when and only when S obtains, and M results when and only when C is tokened, so the link between C and the senses should be the one that maximizes the correlation between C and S. Because the animal labors under constraints, this correlation cannot be perfect, but the animal should seek as good a correlation as it can manage.

The idea that valuable or virtuous cognition consists in a high degree of correlation between thoughts and the states of the world that the thoughts are about is familiar and well-explored in recent philosophy. This is the central idea of reliabilism in epistemology.[2] The appeal to reliable correlation has also spread from epistemology to the philosophy of mind, and it has become popular to claim that the semantic properties of internal representations depend, at least in part, on some kind of reliable correlation between thought and the world.[3] The relevance of mind/world correlation to various philosophical problems is not the chief concern of this paper. Rather, the aim is to develop a reasonably precise theory of what a valuable mind/world correlation is.

---

[2] See D. M. Armstrong, *Belief, Truth, and Knowledge* (New York: Cambridge, 1973); Alvin I. Goldman, "Discrimination and Perceptual Knowledge," this JOURNAL, LXXIII, 11 (November 1976): 771–91, and *Epistemology and Cognition* (Cambridge: Harvard, 1986); Fred Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT, 1981); and Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard, 1981).

[3] See Dretske, *op. cit;* Robert Stalnaker, *Inquiry* (Cambridge: MIT, 1984); and Jerry A. Fodor, *Psychosemantics* (Cambridge: MIT, 1987).

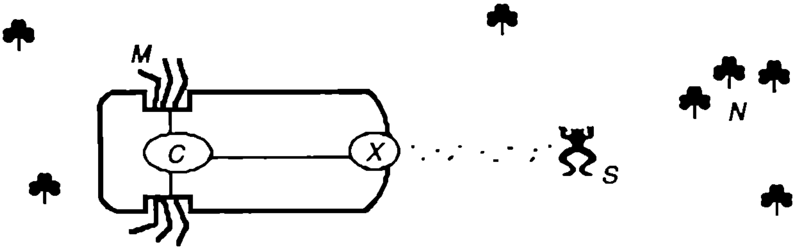


Figure 1

The first essential point is that there are two distinct kinds or "directions" of correlation to consider, and one of these is far more prominent in philosophical discussion than the other. Following Hartry Field,[4] I shall call the currently favored direction of correlation ***head-world reliability***. A perfect degree of head-world reliability exists when an inner state is tokened only if a certain external condition obtains. It is not required that the inner state always be produced by the external condition. So *C* is sufficient, but perhaps not necessary, for *S*. The second direction of correlation is simply the converse. A perfect degree of *world-head reliability* obtains if *C* is always tokened when *S* obtains, whether or not *C* obtains at other times as well. It is not hard to remember which direction is which, as Field's labels follow the direction of a conditional with an arrow '→'. Extreme head-world reliability obtains when a state of the head implies a state of the world; strictly, it is head → world reliability. The converse is world → head reliability. It does confuse matters slightly when we leave the extreme case of perfect correlation, and express the degree of correlation as a conditional probability. Then a high degree of head-world reliability is a high value of $Pr(S|C)$, and a high degree of world-head reliability is a high value of $Pr(C|S)$. But many people find that thinking of Field's labels as abbreviated expressions using an arrow is a useful mnemonic.

It is a central contention of this paper that head-world reliability has been wrongly elevated to a central position in explanations of virtuous cognition. This is explicit in the work of Dretske, Armstrong, and Goldman. Dretske[5] has built a detailed theory of knowledge, perception, and belief content around a notion of correlation that is purely a matter of head-world reliability. Armstrong's 1973 account of knowledge and rational belief makes similar demands. In Goldman's magnum opus, "reliability" is central to his account of good cognition, and for Goldman "reliability" is simply head-world reliability.[6] Goldman also sometimes discusses "power," which is admitted as another valuable property of cognition, and one that must be traded off against reliability. Power is a relative of world-head reliability, but they are not the same thing. Although this is not the place for detailed scholarship on the matter, I think it is clear that these are not isolated cases, that head-world reliability is central to many analytic and naturalistic conceptions of good, responsible, duly cautious cognition.[7]

---

[4] " 'Narrow' Aspects of Intentionality and the Information-theoretic Approach to Content," in *Information, Semantics and Epistemology*, E. Villanueva, ed (New York: Blackwell, 1990).

[5] See *op. cit.* and also his *Explaining Behavior* (Cambridge: MIT, 1988).

World-head reliability has had a little more attention in the philosophy of mind. Although Dretske seeks to explain meaning with head-world reliability, the versions of indicator semantics proposed by Stalnaker (*op. cit.*) and Fodor (*op. cit.*) require both kinds of correlation. Nozick (*op. cit.*) has a similar two-way position in epistemology. Although this is a step in the right direction, I shall argue, the accounts given by Fodor, Stalnaker, and Nozick neglect or distort the most important fact about the relations between the two kinds of reliability; for they apparently require that thinkers simply bring their thoughts into both kinds of correlation with the world. But the critical fact here is the impossibility of pursuing both kinds of reliability simultaneously. Except in the most epistemically benign and friendly environments, the two kinds of reliability cannot be bought as a package deal; rather, one must be traded off against the other. This is seen readily in extreme cases; one way to achieve excellent head-world reliability is to believe almost nothing, but this sacrifices almost all of one's world-head reliability. World-head reliability can be pursued by believing everything that enters one's head, but this also leads to many false beliefs. The remainder of this paper will present a quantitative theory of the reliability trade off, and describe how to strike the balance optimally in particular situations.

I. OPTIMALITY AND RELIABILITY

My discussion of virtuous correlations between mind and world is based on signal detection theory. Signal detection theory originated in the 1950s, combining statistical decision theory with electrical engineering to give a general description of both ideal and real signal-receiving devices. It has been used extensively in psychophysics, and has more recently been discovered by foraging theory in biology.[8] I shall introduce the apparatus in an anthropomorphic, biologically unrealistic way at first, and then show how a core directive of the theory can be mechanized.

---

[6] See his *Epistemology and Cognition*, esp. chs 1–6.

[7] The work of Isaac Levi is an important exception; see his *Gambling with Truth* (New York: Knopf, 1967). In a much longer work (in preparation) I discuss in detail the place of the two kinds of reliability in contemporary epistemology and philosophy of science.

Suppose you are a carnivore, hunting in long grass, and you see a movement in one spot in this grass. This might be something edible creating the disturbance, or it might be just the wind. The decision problem is: Should you bound aggressively to that spot or not?

Assume first that your sensory input can be represented with a single variable X. A high value of X corresponds to the input from a large disturbance of the grass. The mathematics are simpler if we assume that there is only a finite number of values this variable can take. You also know a function relating various values of input X probabilistically to the presence of prey (signal), and a corresponding function for noise. That is, you know how likely each value of X is, given the presence of prey, and you know how likely that same input is to occur when there is no prey there. So what you know can be represented as in figure 2.

Earlier I said that the model would apply to cases where the animal is operating under constraints imposed by imperfect senses and an uncooperative environment. Figure 2 represents this fact in the overlap of the two functions. Degrees of input between 7 and 11, inclusive, can be produced by both signal and by noise. It is important not to read the value of X in the middle of the overlap ($X = 9$) as an input equally likely to have resulted from signal or noise; this depends on the unconditional, or prior, probabilities of signal and noise, as well as on the information in figure 2. In this discussion, the prior probability of signal, $Pr(S)$, will always be 0.2, so the prior for noise, $Pr(N)$ will be 0.8.[9]

A central measure in signal-detection theory is the ***likelihood ratio***, $l(x)$, for an input. Let us avoid a profusion of subscripts in what

---

[8] A standard reference is D. M Green and J. A. Swets, *Signal Detection and Human Observers* (New York: Wiley, 1966). For psychophysical applications, see J. A Swets, ed., *Signal Detection and Recognition by Human Observers: Contemporary Readings* (New York: Wiley, 1964). For applications in biology, see J. A. Staddon, *Adaptive Behavior and Learning* (New York: Cambridge, 1983); and D. Stephens and J. Krebs, *Foraging Theory* (Princeton: University Press, 1986).

[9] The units of probability on the vertical axis are 0.0277 recurring.

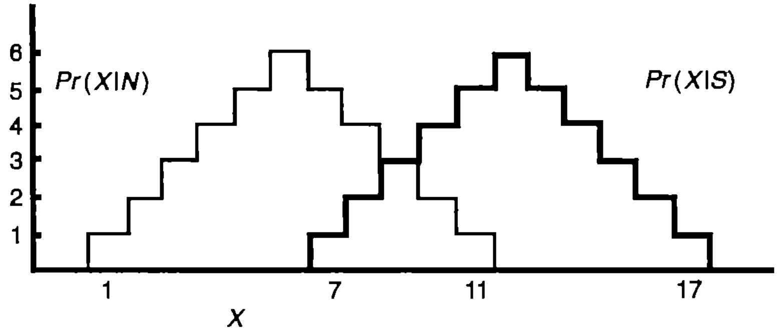


Figure 2

follows by reading an unembellished lower case 'x' as standing for some particular value of the variable X, some x_1.

$$\text{(1)} \qquad \mathrm{l}(x) = \frac{Pr(x \mid S)}{Pr(x \mid N)}$$

It is a consequence of my discrete example that the likelihood ratio is undefined for values of X to the far right, where the value of $Pr(X \mid N)$ is zero. For these values of X, which do not matter much in this discussion, $\mathrm{l}(x)$ will be set artificially at infinity.

With the information in figure 2, and the prior probabilities for signal and noise, we can use Bayes's theorem to tell us what the probability is, given any input x, that there is prey in the grass.

$$\text{(2)} \qquad Pr(S \mid x) = \frac{Pr(x \mid S)Pr(S)}{Pr(x)}$$

Similarly,

$$\text{(3)} \qquad Pr(N \mid x) = \frac{Pr(x \mid N)Pr(N)}{Pr(x)}$$

The unconditional probability of x, $Pr(x)$, is equal to

$$Pr(x \mid S)Pr(S) + Pr(x \mid N)Pr(N),$$

but this is not important for the moment, as $Pr(x)$ cancels when from (2) and (3) we arrive at

$$\text{(4)} \qquad \frac{Pr(S \mid x)}{Pr(N \mid x)} = \frac{Pr(x \mid S)Pr(S)}{Pr(x \mid N)Pr(N)}$$

Readers who find this opaque need not be concerned, as formulas (2) through (8) are only important in their justification of decision rule (9), which we will come to in a moment. Much of the algebra can be skipped, and rule (9) taken on trust, if preferred.

Figure 2 and the prior probabilities of signal and noise tell you a lot about the likely states of the world behind your inputs, but none of this information tells us what you should *do*. To decide what to do, you must take into account the various ways of being right and being wrong, and their consequences. There are in fact four possible outcomes:

> (i) You decide there is prey, and there is. This is called a ***hit***. You attack and get fed, and the value of a hit, or $V(H)$, here is 16 units, let us say.
> (ii) You decide there is prey, but there is none. This is a ***false alarm***.[10] You waste energy attacking, and maybe scare away other animals. The cost of a false alarm $V(FA) = -2$.
> (iii) You decide there is no prey, and indeed there is not. This is a ***correct rejection***. The value of a correct rejection $V(CR) = 0$.
> (iv) You decide there is no prey, but there is. You miss out. This is called a ***miss***. You have not actually spent anything, so the value of a miss here $V(M) = 0$.

Some might think a miss should have a definite cost. In fact it does not matter, for we are envisaging a limited number of opportunities for hits, and a competitive process between different decision rules. Misses are then effectively penalized. These pay-offs can be represented in a simple matrix.

The Foraging Matrix

$$\begin{array}{lll} & S & N \\ \text{yes} & V(H) = 16 & V(FA) = -2 \\ \text{no} & V(M) = 0 & V(CR) = 0 \end{array}$$

S = prey; N = no prey
yes = attack; no = ignore

The expected pay-off from deciding to attack will then be:

$$Exp(yes|x) = V(H)Pr(S|x) + V(FA)Pr(N|x) \tag{5}$$

and the expectation from deciding no:

$$Exp(no|x) = V(CR)Pr(N|x) + V(M)Pr(S|x) \tag{6}$$

---

[10] From the standpoint of the theory of statistical testing, if noise is a "null hypothesis," a false alarm is a type 1 error and a miss is a type 2 error. The relations between signal detection theory and other aspects of statistical theory are discussed in the longer work referred to in note 7.

Assuming that you are interested in simply maximizing your expected pay-offs, case by case, you should say yes if and only if:

$$(7)\quad V(H)Pr(S|x) + V(FA)Pr(N|x) \geq V(CR)Pr(N|x) + C(M)Pr(S|x)$$

This rearranges to give:

$$(8)\qquad \frac{Pr(S|x)}{Pr(N|x)} \geq \frac{V(CR) - V(FA)}{V(H) - V(M)}$$

Substituting (4) into (8) gives, with some rearranging:

$$(9)\qquad \frac{Pr(x|S)}{Pr(x|N)} \geq \frac{Pr(N)}{Pr(S)} \times \frac{V(CR) - V(FA)}{V(H) - V(M)}$$

The expression on the left of the inequality we know as the likelihood ratio l(x), and the expression on the right is known as β. Within β, the quantity $V(CR) - V(FA)$ is the ***importance of noise***. It combines the outcomes of the different actions you might perform as a response to mere noise. The quantity $V(H) - V(M)$ is the ***importance of signal;*** it combines the pay-offs for actions in the presence of signal. The quantities l(x), β, and, within β, the importance of signal and noise, are the important parts of signal detection theory for us, and (9) is the important formula. It directs that you say yes, attack, if the likelihood ratio, calculated from the particular input you are now receiving and its relation to the probability functions for signal and noise, is greater than or equal to β, which is determined by the prior odds of signal and the pay-off matrix. (Strictly, it does not matter what you do when l(x) = β.) In this particular case:

$$\begin{aligned} l(x) &\geq \frac{.8}{.2} \times \frac{0 - (-2)}{16 - (0)} \\ &\geq 1/2 \end{aligned}$$

The value of l(x) beyond which you should say yes determines in turn the critical value of X, the intensity of sensory input that you should consider sufficient to convince you to act. This is shown in figure 3. The critical value of X is labeled x_c. For any value of X at or higher than x_c, you should say yes.

We can see that in this graph the ratio between $Pr(X|S)$ and $Pr(X|N)$, where the value of X is x_c, is ½. As the critical value of l(x) delivered by formula (9) increases, the cutoff x_c moves to the right, and it moves to the left as l(x) decreases.

We are now in a position to answer the questions that began this paper. We asked if there is a general principle describing the opti-

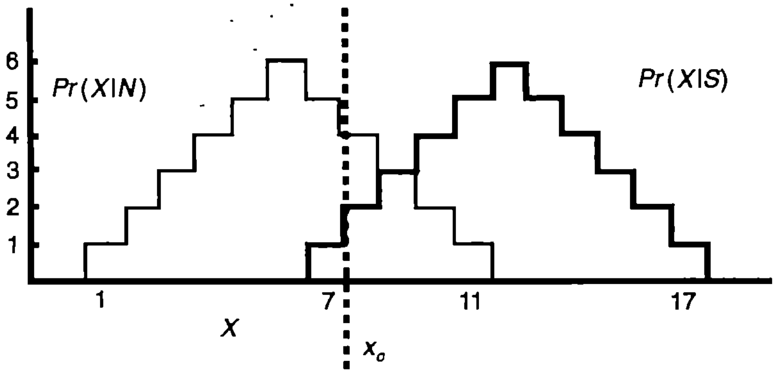


Figure 3

mal way for an animal to have an action-guiding inner state wired to its senses in order to achieve a desirable correlation between its thoughts and the world. Formula (9) is just such a principle. Given that *M* is caused by *C*, and the animal wants to produce *M* if and only if $l(x) \geq \beta$, the optimal *C* is a state tokened if and only if $l(x) \geq \beta$. So the animal in figure 1 is optimal if the link between *X* and *C* is such that *C* only fires if $l(x) \geq \beta$.

Crucially, the optimal animal does not need to represent to itself any of the algebra through which we have gone. It just has to have connected to *M* a state *C* that, as a matter of physical contingency, fires when the animal's sensory input is such that $l(x) \geq \beta$. It does not know that these are the inputs such that $l(x) \geq \beta$; it just happens to be physically tuned to that set of inputs.

Two objections can be raised and answered at this point to clarify the proposal. First, how much depends on the assumption that *X*, the sensory input, varies in one dimension only? Secondly, the optimal animal described above apparently does not really need *C* at all. The signal detection apparatus only describes the relations between *S*, *X*, and *M*. *C* is an epiphenomenal way station on the road to *M*. These objections can be answered simultaneously. It is not essential that the input *X* be one-dimensional; it can be as multidimensional as you please. It can extend over time, and can combine the states of a number of sense modalities along with memory. All that matters is that it make sense to assign a likelihood ratio to each possible total input. It must be possible to compare the probability of receiving that input as a consequence of signal and as a consequence of noise. Further, once we leave the idealized one-dimensional case, *C* does not look so unimportant in the architecture. C, the inner state, is then the place where the different components of the input are combined.

Although formula (9) describes the optimal wiring between the senses, described by X, and the inner state C, it has not yet told us about the relations between C and S. We have not been told how an optimal inner state is correlated with the world. But this is easy to calculate from the information we have, for formula (9) is a description of how best to strike a balance between the competing demands of head-world and world-head reliability. Once formula (9) has established a cutoff point on the X axis, simple algrebra tells us the head-world and world-head reliability values that result. The fastest way to do the calculation is to note first that the world-head reliability of C, $Pr(C|S)$ is simply the proportion of the area under the $Pr(X|S)$ function, to the right hand side of x_c. This area is the probability that S will produce a value of X that the animal will regard as sufficient to trigger C. In this case, $Pr(C|S)$ is about 0.972.

Once we have this figure, the head-world probability falls out of Bayes's theorem (formula (2) above, with C substituted for x). We first work out the unconditional probability of C, as the overall probability that a value of X will be produced, by signal or noise, which exceeds the cutoff x_c. This is done by summing the value of $Pr(x|S)Pr(S) + Pr(x|N)Pr(N)$, for each value of X over x_c. When this figure is divided into the product of the world-head reliability and the prior for signal, the head-world reliability is the result. In this case it is about 0.467. The animal has sacrificed head-world reliability in order to obtain a very high value of world-head reliability; it is tolerating a large number of false alarms in order to achieve as many hits as it can.

Generally, the further the cutoff point is shifted to the left, the more head-world reliability is being sold off in favor of world-head reliability. It is easy to see how this trade is driven by formula (9). When the cutoff x_c is shifted to the left, this is because a low value of $l(x)$ is being demanded by β in formula (9). Ignoring perturbations in the prior probability of signal, a low value of β in formula (9) will be due to a large "importance of signal" in the denominator, compared to the "importance of noise" in the numerator. Perhaps hits are very valuable, false alarms are very cheap, misses are very expensive, or some combination of these holds. On the other hand, as false alarms become expensive, or the importance of noise increases in other ways, the value of $l(x)$ demanded by formula (9) will rise. Then the cutoff x_c will be forced to the right, and head-world reliability will increase at the expense of world-head reliability. In sum, as the importance of signal increases relative to noise, it becomes rational to trade head-world reliability for world-head reliability. As the importance of noise increases relative to signal, it becomes rational to trade world-head for head-world reliability.

So far we have only considered an example involving foraging for food. By shifting our attention to other goals, other ways of making the trade off can be illustrated quantitatively.

II. LOVE AND DEATH

The next case will be one in which the basic form of the trade off is the same, but for different reasons. This is the problem of avoiding one's own predators. A pay-off matrix for interactions with predators might be as follows.

Predation Matrix

S = predator in the vicinity; N = noise
yes = flight or concealment response; no = no behavior

Here the single most important case is the miss; if you fail to detect a predator, you risk injury and death. Hits and false alarms cost −2 units alike. Correct rejections have no consequences. Leaving the prior probabilities of S and N as they were in the previous case, the value of β becomes 0.2 (see figure 4).

In this case the average head-world reliability of C drops to about 0.375. World-head reliability, on the other hand, is forced all the way to 1. Because of the cost of a miss, the importance of signal is so high in this case that every possible degree of world-head reliability is purchased.

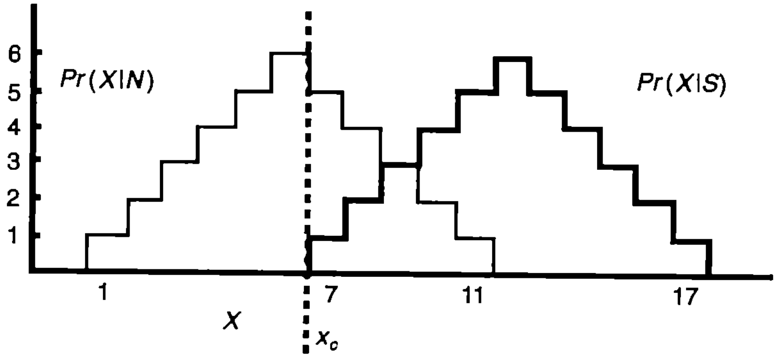


Figure 4

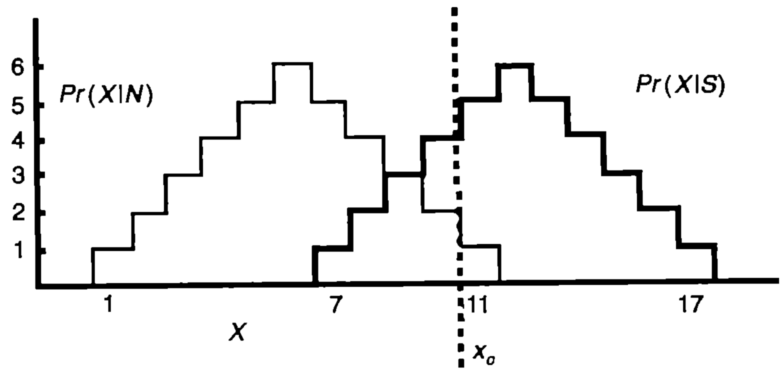


Figure 5

Finally, we might look at a case where the situation is reversed, and head-world reliability should be sought. A suitably ecological case might be one's relations with the opposite sex. We assume the animal is female, and in an environment where there are many males around who are members of different but closely related species to her. Her decision problem is the detection of true conspecifics with which to mate. Here is a matrix:

The Mate Choice Matrix

	S	N
yes	40	−50
no	0	0

S = conspecific male; N = male of another species
yes = mate; no = don't mate

In this situation both hits and false alarms are important. A hit results in pregnancy that costs −50 units, but also in offspring that is worth 90 units. A false alarm is getting pregnant to a male of a different species. The result will be inviable or infertile offspring, so the animal bears all the costs of pregnancy with no reward. Correct rejections and misses have no consequences; there is no shortage of males. The importance of noise is higher than in the previous examples, though the importance of signal is significant as well. Leaving the prior probability of signal at 0.2, the value of β is 5 (see figure 5).

The high value of β pushes x_c to the right and, as a consequence, the head-world reliability of C is now about 0.867. The animal has paid for this with a world-head reliability of 0.722. It is tolerating more misses in its attempt to avoid false alarms.[11] Only in this last case does the optimal animal have the kind of head-world reliability that many philosophers have prescribed for good cognition. If a loosely reliabilist perspective on knowledge is assumed for a moment, then only in this last case might the optimal animal *know* that *S* obtains, when it tokens *C*. As this is largely a consequence of the pay-off matrix operating, a rough slogan suggests itself: the importance of knowledge is the importance of noise.

III. APPLICATIONS

Philosophers have focused on head-world reliability as an important raw ingredient in analyses of justification, meaning, and related concepts. But signal detection theory puts head-world reliability in its place, as a commodity to be sought ardently in some situations, but to be traded off in favor of world-head reliability in others. There are many philosophical projects that might be overhauled in the light of this fact.

In competition with the reliability-based approaches of Dretske, Stalnaker, Fodor, and others, a number of philosophers have recently proposed analyses of the content of mental representation which rely entirely on the biological functions of the representations and the devices that make use of them.[12] A version of this idea can be developed within the signal detection framework. Suppose that the history of an animal's interactions with some state of the world *S* has produced a state *C*, whose location in the animal's wiring is as close to optimal as could be managed in the circumstances. That is, some selective force has led the animal to approximate the optimal trade off between the two kinds of reliability in the wiring of *C*. If *C* has been wired as a detector of *S* in this way, it might have the function to represent *S*, and hence have *S* as its content. This is certainly too crude as it stands—we should have stopped expecting to hear four-line theories of meaning by now. But it suggests a way in which the teleological/functional view of meaning can be made more precise than it is presently, or perhaps a way to steer a middle course between reliabilist and teleological approaches, Dretske, in his recent work, is already moving closer to such a middle course, and is held back only by a preoccupation with head-world reliability.[13]

---

[11] Interestingly, it has not paid that much in world-head reliability to generate this very high head-world value. This fact is also discussed in my longer treatment, along with other features of the signal detection theory apparatus that I have simplified or skipped in this exposition.

[12] See Ruth G. Millikan, *Language, Thought and Other Biological Categories* (Cambridge: MIT, 1984), and "Biosemantics," this JOURNAL, LXXXVI, 6 (June 1989): 281–97, David Papineau, *Reality and Representation* (New York: Blackwell, 1987); and Mohan Matthen, "Biological Functions and Perceptual Content," this JOURNAL, LXXXV, 1 (January 1988): 5–27.

Whether the explanation of meaning can benefit from signal detection theory or not, the theory has important consequences for the program of naturalistic epistemology. Most immediately, the signal detection perspective issues a challenge to those philosophers who claim special status for head-world reliability. This challenge has obvious similarities to pragmatist rebellions against orthodox epistemic values, sharing an emphasis on action and the practical consequences of cognition. Note, however, that while I assume a biological, behavior-linked determination of the pay-off matrices in this discussion, someone who assigned pay-offs on the basis of pure "epistemic utility" could make use of the same formal apparatus. Further, in contrast to many central strands of pragmatist thought, correspondence truth has a definite place in the signal detection approach. For, if C is regarded as representing S, hits and correct rejections are truths, while misses and false alarms are errors. So truth remains a definite goal of cognition, though the truth-linked virtue of "reliability" is taken apart and overhauled.

PETER GODFREY-SMITH

Stanford University

---

[13] See especially his *Explaining Behavior*.

BOOK REVIEWS

Mental Content. COLIN MCGINN. New York: Blackwell, 1989. vi, 215 p. Cloth $43.95, paper $19.95.

According to *externalism*, a thinker's environment contributes to the contents of many of her propositional attitudes. Colin McGinn's book, *Mental Content*, is a welcome contribution to the assessment of the scope, limits, and implications of externalism. Granted, the contrast between facts internal to, and facts external to, a person or individual might seem somewhat elusive. Nonetheless, if externalism is correct, then not all of a person's mental properties are (intrinsic or monadic) properties of her brain. Nor do all of a person's mental properties supervene on physical properties of her brain. Some at least involve relations between an individual's brain and her nonmental environment. Admittedly, this version of externalism depends on an indiscriminate concept of a person's environment. Early on (2, n5) McGinn rightfully (in my opinion) distinguishes the contribution the "nonmental world" might make to the content of a person's mind ("externalism" proper) from the contribution possibly made to the content of a person's mind by what other minds in her community may think ("anti-individualism").

If correct, externalism has many far-reaching philosophical consequences for the epistemology of mental states, first-person authority (about the content of a person's own mind), personal identity, skepticism, the nature of mental properties, physicalism, mental causation, and the role of content in scientific psychology. Is it correct?

McGinn diagnoses the major source of resistance to externalism in *substantialism*, the view that the mind is a substance (14–30, 100–3): "if externalism were true of the mind, then it could not have the substantial nature of the body or brain" (17, 23, 116). So consider, e.g., personal identity. If some of a person's mental properties change when his environment changes, then it might seem that his personal identity might concomitantly be affected. This is a consequence McGinn deflects by distinguishing what confers upon a person his mind from what makes a person the person he is (3, 24–6, 46): whereas a person is a substance, a mind is not (25).

At the core of the book is McGinn's distinction between *strong* and *weak* externalism (7–9). According to strong externalism, unless a thinker causally interacts with a particular spatiotemporal item present in her nonmental *environment*, she could not entertain a

0022-362X/91/8812/723–8

thought with a particular content. Prime candidates for strong externalism are contents of thoughts about particular spatiotemporal material objects and primitive (nondefinable) natural-kind concepts. According to weak externalism, unless property *P* is instantiated by a particular item somewhere or other in the nonmental *world,* a thinker could not entertain a thought, one constituent of which is a concept denoting or expressing *P*. Candidates for weak externalism are abstract arithmetical or geometrical properties and perceptual content.

McGinn's book is a critique of "total strong externalism." The crux of his argument is that property realism and a teleological account of content, both of which McGinn favors, conspire to entail a *weak* externalist view of some contents.

What is relevant to, e.g., perceptual content is less its causes than its behavioral effects (58–99): while I have the visual experience as of a square shaped building prompted by my perceiving a square building in the actual world, my twin may, the weak externalist claims, have an indistinguishable visual experience prompted by a round building transmitted by warped light rays on twin earth. Guided by the same perceptual content, we may both exhibit the same squarewise driving behavior. McGinn takes the strong externalist verdict of a mismatch between my twin's perceptual content and his behavior on twin earth as a refutation of total strong externalism (65–6). But now, if some contents are weakly external, then total internalism is false and so is total strong externalism. I shall argue that weak externalism generates a dilemma between naturalism and property realism.

Mental Content has three parts. In the first part ("The Location of Content," which occupies over half the book), McGinn does two things: he lays out his metaphysics and probes the applicability of (weak and strong) externalism to a variety of different mental contents. As the second part of the book, "The Utility of Content," makes clear, McGinn's outlook is broadly naturalistic and falls squarely in the recent tradition of teleological accounts of content[1] which, he says, he "was slow to come round to" (143). In the third part, he espouses the mental-model view of what he variously calls the "basis," the "ground," or the "mechanism" of content.[2]

---

[1] See Fred Dretske, "Misrepresentation," in R. Bogdan, ed., *Belief* (New York: Oxford, 1986); Jerry A. Fodor, *A Theory of Content* (Cambridge: MIT, 1990); Ruth G. Millikan, *Language, Thought and Other Biological Categories* (Cambridge: MIT, 1984); David Papineau, *Reality and Representation* (New York: Blackwell, 1987).

[2] A view he ascribes to Cambridge psychologist Kenneth Craik and which has been recently rehabilitated by Cambridge psychologist Philip Johnson-Laird, *Mental Models* (New York: Cambridge, 1983).

There is a connection between the teleological account of content advocated in part II and the mental-model view of the "basis" (or "ground") of content advocated in part III. In the latter, McGinn addresses what he takes to be an "engineering" or a "design" task —a task he calls "psychotectonics." Roughly, the task is to build a machine that "invites the assignment of propositional items, not to say what these assigned items themselves consist in" (170–2). McGinn thinks there are two major competing views that may claim to do the job: Jerry Fodor's[3] sententialist hypothesis of a language of thought and the mental-model view. McGinn believes (182) that, once the confusion between the "ground" or "mechanism" of content and the (propositional) "object" or "target" is cleared up, then the mental-model picture can be seen to be the correct view of the former.

His reason lies in the analogy between the role of (abstract) propositions as *indices* in assigning contents to mental states and the role numbers play in scientific measurement as indices of physical magnitudes (107, 139–43, 182–3, 187). One anti-Fodorian consequence of the indexing analogy between contents and numbers is that "logical structure" need no more be "literally *in* the head" than "indices need . . . in general be *in* what they index" (183). Unlike assignment of numbers in measurement, however, assignment of (propositional) content to mental states has a teleological point (142): the content of a representational state is linked to the function the state has in delivering valuable information for the benefit of the organism whose state it is.[4] Both the analogy and the disanalogy between numerical indices in measurement and propositional indices in content assignment loom large in McGinn's project since he depends on them for his argument that content is suitable for scientific psychology.

Since Hilary Putnam's[5] seminal thought experiment, philosophers have disagreed about the suitability of content for a naturalistic approach to the mind or for scientific psychology. I found insightful McGinn's criticism of Paul Churchland's[6] charge that contentful folk psychology is a degenerate stagnant research program (123–6).

---

[3] *The Language of Thought* (New York: Crowell, 1975), and *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge: MIT, 1987).

[4] McGinn suggests that computers, which are purely syntactic engines, lack altogether "original teleology," "affective or conative states," and intentionality (155): "intentionality begins with wanting, . . . not thinking" (155, n56).

[5] "The Meaning of 'Meaning'," in *Philosophical Papers, II: Mind, Language and Reality* (New York: Cambridge, 1975), pp. 215–71.

[6] "Eliminative Materialism and Propositional Attitudes," this JOURNAL, LXXVIII, 2 (February 1981): 67–90.

So did I find his discussion of Steven Stich's[7] castigation of content and defense of a syntactic scientific psychology, and of Putnam's dichotomy between algorithmic, computational, subpersonal scientific psychology and informal, holistic, personal, normative content-based interpretation theory (127–32). McGinn's conclusion is "not . . . to banish content from the science of psychology" (132, 156), but to recognize that biology, not physics, ought to be the proper paradigm for scientific psychology. One assumption used to reach the conclusion is that subpersonal content of the kind ascribed either in low-level human visual information processing or to nonhuman animals, though no less content than the content of propositional attitudes, is less subject to hermeneutic perplexities.

Whereas he sees (rightly in my opinion) no tension between externalism and a physicalist ontology (103–5), McGinn takes very seriously the puzzle externalism generates for mental causation, i.e., the threat that the mental be epiphenomenal. The puzzle arises from the tension between three views: the externalist view that some content is relational (or extrinsic to an individual's brain); the view that, however relational, content may be a causally efficacious property; and the view (which McGinn espouses) that all causation is a *local* process (133–7). Pondering over this tension, many philosophers have recently worried whether physical properties of an individual's brain do not screen off or preempt mental properties from causal efficacy or causal relevance.[8]

McGinn "concedes" the epiphenomenalism of the mental and argues that the concession will seem fatal only to philosophers in the grip of physics worship. On his view, genuine causal explanations are physical explanations: they must refer to physical properties of mechanisms. "Content cannot be saved on the strength of what it takes for the causal wheels to turn" (149). He insists, however, "that there is more to explanation than simply citation of causes" (144). His explanatory vindication of content is threefold. First, from the indexing analogy between numbers and contents, he argues that contents are neither more nor less causally efficacious, hence neither less nor more explanatorily respectable, than numbers (141). Second, he argues, in biology, not only does the information supplied by "teleological taxonomies" differ from, and complement, the information supplied by "causal taxonomies" (145–6), but the

---

[7] *From Folk Psychology to Cognitive Science* (Cambridge: MIT, 1983), and Putnam, "Computational Psychology and Interpretation Theory," in *Philosophical Papers, III: Realism and Reason* (New York: Cambridge, 1983)

[8] For a sample, see J. Tomberlin, ed., *Philosophical Perspectives*, III: *Philosophy of Mind and Action Theory* (Atascadero, CA: Ridgeview, 1989).

"biological function of a trait, mental or bodily, cannot be revealed by examination of the causal mechanisms" (149). Finally, from the teleological account of content, he argues that content is neither better nor worse off than the biological function of a trait.

Let me now go back to "the crux" of McGinn's argument against total strong externalism. One obvious rationale for the distinction between strong and weak externalism is that it provides a neat alternative to the distinction between broad and narrow conceptual contents.[9] Natural-kind concepts are strongly external. The perceptual content of a sensory experience is weakly external.[10] Therefore, the conceptual content of my water beliefs differs from the conceptual content of my twin's twater beliefs. But the *perceptual* content of my experience triggered by my perception of a glass of water and the *perceptual* content of my twin's experience when he perceives a glass of water are indistinguishable—an assumption clearly made in Putnam's seminal thought experiment.[11]

Why does the teleological account of content support *weak* externalism? The reason is that "contentful states can in principle have a function that is specified by reference to a property that is not actually instantiated in the organism's causally impinging environment . . . a function could in principle exist without that function ever actually being fulfilled" (67). Why does property realism support *weak* externalism, too? The reason is that, on McGinn's view, "properties are mind-independent and objective" (116, n130) and "we simply have no other conception of, e.g., the concept *square* than as something that denotes the property of being square" (41, n57). It is, McGinn believes, a condition of adequacy on any theory of content that it solves what he calls (11–2) "the matching problem": How do concepts match (or "latch onto") properties of things? On McGinn's "naive realist" view, mind-independent objective properties have a dual role or may enter into two distinct relations with objects: they may be instantiated by, or predicated of, objects. If an object instantiates a property and if an individual's thought predicates the latter of the former, then the thought is true.

Here is the dilemma an advocate of weak externalism faces, I think: either for me correctly to ascribe the property of being square, I must causally interact with square-shaped spatiotemporal particulars in my environment, or not. If the former, strong externalism is vindicated. But McGinn cannot take this option, since on his view the concept of a square shape is typically an "observational" concept (59) to which weak externalism applies as it does to perceptual content. If the latter, then either naturalism is false or property realism is false: either the mind has the nonnatural power to reach out into the world to the property of being square uninstantiated in its environment or being square is a projection from my concept. Suppose "possessing the concept square cannot consist in anything other than standing in some relation (the intentional relation) *to* that property" (41, n57). Unless being square is a *projection* of my concept of a square, my mind, it seems, will not stand in relation to this property if the latter is not instantiated in my environment. But this is to give up property realism.

---

[9] A distinction earlier espoused by McGinn in "The Structure of Content," in A. Woodfield, ed., *Thought and Object* (New York: Oxford, 1982), pp. 206–58.

[10] On the strong externalist view of perceptual content, "aetiology trumps aftermath" (65); on weak externalism, "aftermath trumps aetiology" (66).

[11] "The Meaning of 'Meaning'," p. 224.

PIERRE JACOB

C.N.R.S. Ecole Polytechnique

Cement of Society. JON ELSTER. New York: Cambridge University Press, 1989. 311 p. Cloth $49.50, paper $16.95.

In a career that has resulted in the publication of nine books between 1978 and 1989, Jon Elster has shown a deep commitment to two positions he has defended as mutually consistent, namely, Marxist political conclusions and the individualistic methodologies embraced by microeconomics, rational-choice theory, and certain brands of political theory. While the Marxist commitment is not particularly prominent in *The Cement of Society,* this book not only manifests but is also in large part motivated by Elster's individualism. The book's project is to explain how groups of individuals constituting a society remain together—how they are, as it were, cemented into a group.

As Elster points out, such a project presupposes the view that individuals are conceptually prior to societies: "There are no societies, only individuals who interact with each other. Yet the structure of interaction allows us to identify clusters of individuals who interact more strongly with each other than with people in other clusters" (248). In order to measure the cohesiveness of these clusters, he defines a "coefficient of cohesion," which is the number of transactions between individuals in the area divided by the total number of transactions in which these individuals are involved; and then defines a society as an *area* (not a set of people) that has a local

0022-362X/91/8812/728–38

maximum of cohesiveness (248). Thus, on his view, understanding a cooperative, stable society requires understanding the reasons individuals persist in cooperative interactions in this area. To do this we must, he claims, analyze "two concepts of social order," i.e., "stable, regular, predictable patterns of behavior" and "cooperative behavior" (1). Making liberal (but not uncritical) use of game theory and economics, Elster explores how, in a variety of ways, unpredictability and noncooperative behavior are the acids that corrode the bonds of society, although he also shows how, paradoxically, cooperation and predictability can sometimes precipitate disorder and conflict rather than remedy or forestall them.

Many of the ideas in the book are not new. Elster's talents lie not so much in generating new theories out of whole cloth as in articulating a point of view about particular aspects of the human world he aims to study and then imaginatively gathering together theories, ideas, and methods from a variety of disciplines that bear on the issues he is addressing. After introducing the book with a brief review and analysis of existing theories of cooperation and predictability, he reviews and criticizes recent theories of collective action and bargaining in the first two chapters, thereby providing an analytical framework for later discussions of social interactions. The third chapter analyzes what Elster calls "social norms," which he contends provide a nonrational source of human motivation as important as rational self-interest in understanding both the origination and the disintegration of social cooperation. The remaining chapters formulate explanations, descriptions, and evaluations of bargaining in the context of solving collective-action problems (as an example, Elster explores the collective-bargaining experiences of labor and management in Sweden), and the role social norms play both in solving and preventing the solution of collective-action and collective-bargaining problems.

The book is also a defense of what I shall call "middle-level" theory. Elster eschews both mere history of existing social arrangements and abstract, overarching theory construction: instead he tries for small-scale theoretical explanations of patterns of human interactions, openly dismissing the idea that the present state of social science makes possible a full-blown theory postulating general, law-like regularities about human behavior: "we should concentrate on specifying small and medium-sized mechanisms for human action and interaction—plausible, frequently observed ways in which things happen" (viii). So the lack of a systematic method of inquiry into his subject is deliberate on Elster's part; and while the book often strikes the reader as somewhat chaotic and unstructured, on many occasions it offers genuine insights into the workings of the social world. Moreover, its richness of detail and breadth of coverage pull readers away from their favorite abstract theories and force upon them the messy details of actual social reality, thereby helping to free them from the confines of any one limited theoretical point of view.

Still, this book is often simultaneously insufficiently theoretical and insufficiently concerned with details. The pace of the book is fast—inexplicably fast. Sometimes the reviews of literature or the presentation of problems associated with, say, a particular theory of bargaining or an account of labor-management interaction, are so hasty and lacking in detail that the reader has trouble fully understanding the structure of the theory Elster is discussing, or the position he is attributing to an author, or even his own position. Such "hit and run" philosophy will not sit well with the patient, careful analytic philosopher, who wants precise, accurate, and well-crafted analysis and argumentation. More fundamentally, Elster's intellectual hastiness damages his ability to give penetrating analyses of social interaction.

For example, early in chapter 1 he claims that, by thinking of persons as collectives of time slices, we can assimilate weakness of will to an intrapersonal collective-action problem (19ff). Now, if this is true, and if, as Elster insists, a collective-action problem is normally an *n*-person prisoner's dilemma, then these time-slice persons should be in a prisoner's dilemma with one another. They are not, however: in a PD, each player's favorite outcome is that outcome in which he does not cooperate but the other players do, and his second favorite outcome is universal cooperation; whereas in this "weakness of will" game, the earlier selves prefer both the outcome of unilateral noncooperation and the outcome of universal noncooperation to the outcome of universal cooperation (because their cooperation yields only costs and no benefits for them, even when the later selves cooperate). Oddly, Elster himself admits (23) that the two problems are not "strictly analogous," and the puzzled reader wonders why they were ever compared.

Without stopping to explain, Elster moves on to consider possible definitions of "the collective-action problem," selecting one that defines it as an *n*-person prisoner's dilemma "in which it is better for all if *all* cooperate than if nobody does" (25). In the course of defending that definition, he notes that some writers (myself included) have argued that collective-action problems can have non-PD game-theoretic structures, e.g., chicken games.[1] Yet he never revises his narrow definition to accommodate the existence of these non-PD collective-action problems. And when he discusses the role of social norms in solving collective-action problems in chapter 5, he repeatedly uses examples that are not PDs, although he explicitly introduces them as such. Consider three such examples:

---

[1] See my "Free Rider Problems in the Production of Collective Goods," in *Economics and Philosophy*, III (1987): 245–74; and M. Taylor and H. Ward,

> A group of friends who are cleaning up after a party might actually finish the job faster if some of them relax instead with a drink, but the norm against free-riding might overwhelm considerations of efficiency. Joining the army in wartime is a more substantial example. Those who stay home to work in vital industries may feel that they are violating the norm of fairness. If all who want to join were allowed to do so, the war effort as a whole might suffer. Or consider the Leninist theory of party organization. If organizational efficacy is an increasing function of numbers and discipline, and discipline is a decreasing function of numbers, it could happen that the optimal number of party members falls well short of universal membership. Even if there is normative pressure on workers to join the party and nonjoiners are frowned upon, not all can be allowed to join (190).

In each of these situations, there is a social norm "solving" the problem of cooperation "too well," by encouraging all to contribute when the contributions of only some would be optimal. But in a genuine PD there is no such thing as too much cooperation,[2] because universal cooperation is Pareto superior not only to universal noncooperation, but also to all partially cooperative outcomes. So because there are outcomes of only partial cooperation in Elster's examples which are Pareto superior to the outcome of universal cooperation, they cannot be examples of PDs. Moreover, Elster's presentation of the examples is so sketchy that we can interpret players' preferences in them so as to define several different non-PD game-theoretic structures. If we take the players to prefer contributing to the collective good alone (thereby generating at least some portion of the good, e.g., the communist party, the war effort, the clean house), to not contributing at all (and having none of the good produced), then table II represents the examples. If they do not, then the examples are represented by table I.[3] Neither game is a PD.

"Chickens, Whales and Lumpy Goods: Alternative Models of Public Goods Provisions," *Political Science*, III (1982): 350–70.

[2] One of the drawbacks of the Schelling-style diagrams Elster uses to illustrate these examples is that they are not complicated enough to show the non-PD structure of the games in these examples.

[3] But why not interpret the house-cleaning example as follows?

CCC 2, 2, 2
CCN 4, 4, 5
CNC 4, 5, 4
NCC 5, 4, 4
NNC 3, 3, 1
NCN 3, 1, 3
CNN 1, 3, 3
NNN 0, 0, 0

[Note: '*C*' means "cooperate"; '*N*' means "not cooperate"; the numbers represent each person's cardinal utilities (5 is highest) over outcomes that would result from their selection of actions.]

I.	CCC 3, 3, 3	NNC 2, 2, 0	II.	CCC 3, 3, 3	NNC 2, 2, 1
	NCC 5, 4, 4	NCN 2, 0, 2		CCN 4, 4, 5	NCN 2, 1, 2
	CCN 4, 4, 5	CNN 0, 2, 2		CNC 4, 5, 4	CNN 1, 2, 2
	CNC 4, 5, 4	NNN 1, 1, 1		NCC 5, 4, 4	NNN 0, 0, 0

Getting the game-theoretic structure of these—and other—collective-action problems right is important for a number of reasons. First, it affects our analysis of the impact of social norms on cooperative endeavors: it is because the game-theoretic structure underlying Elster's examples is not an *n*-person PD that the social norms he takes to be applicable to these situations are troubling: they are encouraging a level of cooperation that is not Pareto optimal. Indeed, these norms would *not* be troubling if the games really were PD games, because then the norms would encourage a universal level of cooperation that would be both Pareto optimal and equitable. Second, the correct analysis affects how we should think about solving the problems posed by these games: whereas social norms are a good solution for genuine PD problems, they may be poor solutions for games of this sort unless they can be defined or applied so that they do not encourage too much cooperation. Third, an appreciation of the way collective interests and *some* individual interests coincide can lead one to speculate about the kinds of mechanisms and procedures—in particular, bargaining procedures—that could solve non-PD collective-action problems. Finally, getting the game-theoretic structure right strikes me as important for its own sake: if this is a book that seeks to help us understand the structure of social interaction, then it owes us an analysis that is accurate and precise.

The area of the book that best exemplifies both the strengths and the weaknesses of Elster's analysis is chapter 3, where he attempts to define and explore the idea of a "social norm." The incorporation of social norms, the stock-in-trade of sociologists, in a book seeking to understand social cohesiveness using the machinery of decision theory, is an example of how Elster's breadth of reading and openness of mind enable him to take theoretical tools from different fields and combine them in a refreshing and promising way. On the other hand, his analysis of the nature of social norms, although very interesting and philosophically suggestive, has many problems that would trouble both decision theorists and sociologists.

Elster's definition of social norms has two components: insofar as social norms are *norms*, they are "not outcome-oriented" (98); and insofar as they are *social,* they "must be (a) shared by other people, and (b) partly sustained by their approval and disapproval" (99).

Consider the first component. Whereas hypothetical imperatives of the sort economists give us are "outcome oriented," Elster maintains that social norms (which he calls "non-consequentialist" (101)) are "not outcome-oriented," which he says means "unconditional, or if conditional, . . . not future-oriented" (99). As an example of such a norm he gives us "everyday Kantianism," which is the norm to "cooperate if and only if it would be better for all if all cooperated than if nobody did" (123) and which he is careful to note is not really identical to anything in Kant's moral theory:

> In one sense, this norm is outcome-oriented since it refers to the outcome that would be realized if everybody acted in a certain way. In another sense—the sense that is relevant here—the norm disregards outcomes, since it does not enjoin an individual to consider the outcome of *his* action. The norm is conditional upon hypothetical outcomes, not actual ones (123).

But what does it mean to say that the norm is 'conditional upon hypothetical outcomes'? And what is "the relevant sense" of 'outcome-oriented' that makes it inappropriate to say, in *that* sense, that the Kantian norm is outcome-oriented?

Let me propose the following interpretation of what Elster might mean here. Consider that the Kantian norm contains within it directions for how to ascertain when to cooperate, i.e., one cooperates when it would be better for all if all cooperated rather than if no one did. Call these directions the *substance* of the norm. The substance of this Kantian norm makes reference to hypothetical, albeit not actual, outcomes, and in that way the norm is "outcome-oriented." Norms also have, however, what I shall call *authoritative force.* As I shall discuss more fully later, this force is *not* the same as the psychological motivation that someone might have for following the norm: instead *it is the reason why following the command would be justified, correct, appropriate.* The authoritative force of a hypothetical imperative, or indeed any consequentialist command, derives from the assumption that the outcome that the action being commanded would effect is a good, and thus desired or desirable. Even if a nonconsequentialist norm might direct one to take into account the consequences of one's actions in hypothetical circumstances, however, these make-believe consequences are not what make performance of the action right. Instead, such a norm is supposed to have an internal, irresistible authority, which has been called "categorical" (and which many have charged is too mysterious to be taken seriously).

If this interpretation is right, Elster is attributing nonconsequentialist authority to social norms when he says they are not outcome-oriented in a "certain sense," and when he persistently calls them "non-consequentialist." Yet, since he never uses the term 'categorical' to describe social norms, one wonders whether he really holds this view. One's doubts are increased by his insistence that social and moral norms differ, although when he explains how these types of norms differ, he dwells on the difference between *consequentialist* moral norms and social norms that "consist of nonconsequentialist obligations and interdictions, from which permissions can be derived" (101). So how are social norms different from nonconsequentialist moral norms, whose authority is usually called "categorical"? Elster says only that they are very similar (101), and, oddly, does not stress the way the former, but perhaps not the latter, are *social*.

So the reader is never clear what kind of authoritative force Elster wants to attribute to social norms. And since we do not learn this, we also do not learn whether, if he would attribute categorical force to them, he would agree with those (such as J. L. Mackie) who say that this is a "false" authority which human beings unwittingly manufacture and attribute to certain imperatives, or with those (such as Kant) who take it to be an objectively real phenomenon to which human beings are appropriately responsive when they follow certain norms.

Elster's comparison of social norms to private norms only increases the readers' puzzlement over his position on the authority of the former. He claims that "[p]rivate norms, like social norms, are non-outcome-oriented and sustained by feelings of anxiety and guilt" (103), differing from social norms only because "they are not sustained by approval and disapproval of others, since they are not, or not necessarily, shared with others" (103). But when he goes on to give examples of private norms, the examples seem to establish that these norms *are* outcome-oriented both substantively and authoritatively, since they prescribe behavior that, although costly in the short run, is utility-maximizing in the long run. Consider the following example: "I may feel a need to protect myself against an addiction to watching television sports on Sunday simply because there are other things I want to do more, not because my friends would frown on me were they to learn about it" (103). Now, perhaps one who adopted the private norm "don't watch TV" would "feel" as if its force were nonconsequentialist (insofar as this norm would be overriding the strong immediate desire to be a couch potato), but Elster's justification of someone's following the norm is consequentialist: by adopting it one will not develop a habit that will prevent one from filling up a Sunday with more enjoyable (and thus more utility-maximizing) activities. Every other example of a private norm Elster gives has a consequentialist justification of this sort: e.g., norms against drinking, smoking, or overeating; or norms prescribing jogging or saving money (102–3).

I suspect that Elster does not appreciate the consequentialist character of the private norms he describes because he fails to make another distinction, this time between the authority of the norm and its psychological "grip" on the mind. While it is true that, say, the norm to save money may have an internal grip on me that feels nothing like the grip that a counsel of prudence has, if I only subject myself to the former norm because conformity to it will ensure behavior that is utility-maximizing in the long run, then its authoritative force is consequentialist. Indeed, a rational-choice theorist (such as David Gauthier) might be happy to admit that, psychologically, those of us who act from private or public norms often feel as if we are being pressured by something that appears to require us to act without taking into account the consequences or outcome of that action, but they will also interpret these norms as rules dictating behavior that turns out later to be individually or collectively rational, and for that reason justifiably authoritative over our decision making.

So the way in which Elster compares private norms with social norms unintentionally provides rational-choice theorists with the ammunition they need to argue that, *contra* Elster's claim, action from social norms is merely rational, utility-maximizing action in disguise, and thus consequentialist in authority. Moreover, his discussion of social norms plays into rational-choice theorists' hands in another way.[4] At one point he discusses what he calls the "grip" these norms have on us when he says they are partly sustained by the approval and disapproval of other people" (99). But if people follow a norm because they are striving to avoid the expected costs of social sanctions and reap the benefits of social conformity, then their norm following is merely rational behavior all along. Note that a rational-choice theorist who takes this position has a different way of reducing norm-following behavior to rational behavior than the theorist I discussed above. The former theorist argues that, whereas norms have a distinctive motivational effect on people unlike the effect of hypothetical imperatives, nonetheless these norms exist and are justified for consequentialist reasons; that is, following them will be, in the long run, individually or collectively rational. In contrast, this second rational-choice theorist maintains that social norms do not have a special motivational effect on people, but are merely, in the words of Hume,[5] "conventions of men, which create a new motive, when experience has taught us, that human affairs would be conducted much more for mutual advantage, were there certain *symbols* or *signs* instituted, by which we might give each other security of our conduct in any particular incident," and this new motive is avoiding costs associated with the disapproval of other people.

---

[4] Much of the rest of ch. 3 is taken up with challenging the claims of rational-choice theorists that norm following is utility maximizing because the norms themselves are either individually or collectively rational. He charges that theorists who insist on this characterization are simply telling "just so stories" whose plausibility cannot be established. I doubt that committed rational-choice theorists will be convinced by his examples of social norms that he claims make everyone worse off (e.g , norms of etiquette), since his own brief accounts of the costs of such norms are subject to the same charge of being mere "just so stories." Still, his speculations in these pages should encourage those who study the emergence of social norms to be more careful in defining and defending their claim that such norms are collectively rational.

Now, Elster appreciates that someone "might ask whether the existence of sanctions does not make it rational to follow the norm, thus undermining the contrast between rational behavior and norm-guided behavior" (99). He responds by saying that norms are not only enforced by social approval but also by private emotions, such as embarrassment, anxiety, guilt, and shame, so that external sanctions are frequently unnecessary to ensure that people follow social norms (131, 100). But the Humean rational-choice theorist will argue that a person is still utility maximizing if she follows a norm to avoid costly painful *internal* feelings. And this theorist can propose that society inculcates norms in its members at least in part by habituating them such that, once the norms are internalized, they will experience these painful feelings when they break the norms. Note that a Kantian might rescue Elster's argument by maintaining that the costly private emotions often experienced after breaking a norm are merely side effects of the realization that one has contravened an authoritative directive, so that it is the innate authority of the norm, and not the costliness of the emotions triggered by breaking it, which provide one with both the motivation and the reason for keeping the norm. If Elster agrees, then as I have said, this mysterious innate authority must be recognized and explored.

The second component of Elster's definition of social norms defines their sociality with two criteria: first, that the norm is "shared by other people" (99); and, second, that it is partly sustained by other people's approval or disapproval. But as we discussed, if Elster emphasizes the second criterion, his discussion will not only fail to illuminate, but also undermine, the way social norms function "normatively" in the population. In addition, neither criterion, either by itself, or taken together, provides a sufficient condition of a norm's sociality. To see this, consider Elster's example of a *private* norm voluntarily accepted by all the members of many Tahitian villages to involve themselves in the practice of religion. He cites anthropologists who claim they do so "because they remember the savage pagan behavior of their ancestors, the wars and cannibalism (matters which missionary teachings constantly emphasized when they portrayed the salvation from savagery brought by religion), and being afraid of backsliding, use religion to protect themselves from doing evil" (103). Now, since everyone accepts this so-called "private" norm, it certainly seems to be shared by all the members of the village, thereby meeting Elster's first criterion. And we can develop the example further to show how this norm can also meet the second criterion. Suppose a Tahitian villager who embraces this norm confesses to his friends that he has violated it recently, and suppose that, since they have voluntarily accepted the same norm, they express their disapproval of his actions in a way that he dislikes, making it less likely that he will violate the norm again. Nonetheless, even if their disapproval partly explains why he subsequently keeps the norm, that norm would still seem to be a private one insofar as he voluntarily adopted it for reasons of his own. And if private norms can have the feature that they are partly sustained by the disapproval or approval of other people, these social sanctions cannot be what make a social norm *social*. So in what does a norm's sociality consist?

---

[5] *A Treatise of Human Nature,* L. A. Selby-Bigge, ed. (New York: Oxford, 1978; 2nd ed.), bk III, pt. II, sect. V, p. 522.

Is it not the fact that social norms are inculcated in individuals by social institutions? Although the Tahitian norm is shared with others in the sense that they all have adopted the same norm, nonetheless, as Elster himself says, this norm has been "voluntarily accepted as a means to impulse control" by each individual (103). This suggests that a norm is social if its internalization by a person does not come about by a private, voluntary decision but by societal inculcation (which presumably involves the effects social institutions have on the formation of people's character and beliefs). Why does Elster not say this, and explore the social origination of a social norm?[6]

Perhaps he avoids such theorizing because he fears it may not be consistent with his attempt to "define, discuss and defend a theory of social norms within a wholly individualist framework," one which eschews the idea that norms are "supraindividual entities that somehow exist independently of their supports" (105–6). But might it be that, in order to understand their sociality, we *do* have to think of norms as supraindividual entities?

---

[6] But note that, if we can only define a social institution by means of the existence of social norms, this approach cannot work. I am indebted to David Copp for this point.

How well Elster can deny this depends in part on what he believes he is committed to by virtue of embracing an individualistic method. Yet there is no detailed discussion of his method anywhere in the book, and the reader is left confused about what he means to endorse. For example, does his method presuppose that a society or social institution is, in his own words, merely "an area which has a local maximum of cohesiveness"? Yet it seems a mistake to identify a social unit with an *area* or place, and in any case such a characterization fails for social institutions scattered across territory, e.g., the Catholic church or large multinational business corporations. So should we say this method presupposes the simple ontological thesis that only individual human beings exist, and not collectives or properties of collectives? But this will not do, because an individualist could admit the existence of, say, the collective "America" and even attribute properties to it, but define it merely as a certain set or sum of individuals, whose properties are reducible to properties of its individual members. So might it be a reductionist thesis to the effect that social properties can be reduced to individual properties? But why not allow the nonreductionist possibility that social properties—even emergent social properties—can supervene on properties of individuals?

Answers to these questions are not forthcoming in *The Cement of Society*. Of course, Elster would be right to reply that those who invoke an anti-individualist, nonmechanistic approach to social phenomena are even less clear about the details and motivations of their method. Still, one wishes he had defined more carefully his alternative approach to social science. And, more generally, even while appreciating his imaginative assembly of ideas from many disciplines, one wishes he had taken more accurate analytical aim at the social phenomena he attempts to explain.

JEAN HAMPTON

University of Arizona

INDEX TO VOLUME LXXXVIII

INDEX OF SUBJECTS DISCUSSED IN THE JOURNAL

Action, Human
- Hursthouse, Rosalind: Arational Actions. 88 (2): 57–68 '91

Aesthetics
- Alperson, Philip: The Arts of Music. [Abstract.] 88 (10): 555–6 '91
- Kivy, Peter: Is Music an Art? [Abstract.] 88 (10): 544–54 '91

Agency, Social
- Brook, Richard: Agency and Morality. 88 (4): 190–212 '91

Aquinas, Thomas
- Jenkins, John: Aquinas on the Veracity of the Intellect. [Abstract.] 88 (11): 623–32 '91

Arationality
- Hursthouse, Rosalind: Arational Actions. 88 (2): 57–68 '91

Art
- Alperson, Philip: The Arts of Music. [Abstract.] 88 (10): 555–6 '91
- Kivy, Peter: Is Music an Art? [Abstract.] 88 (10): 544–54 '91

Biology
- Bedau, Mark: Can Biological Teleology Be Naturalized? [Abstract.] 88 (11): 647–55 '91
- Matthen, Mohan: Naturalism and Teleology. [Abstract.] 88 (11): 656–7 '91

Chisholm, Roderick M.
- Kaplan, Mark: Epistemology on Holiday. 88 (3): 132–54 '91

Choice
- Hollis, Martin: Penny Pinching and Backward Induction. 88 (9): 473–88 '91

Communication
- Goldman, Alvin I.: Epistemic Paternalism: Communication Control in Law and Society. 88 (3): 113–31 '91

Complexity, Computational
- Barnes, Eric: The Causal History of Computational Activity: Maudlin and Olympia. 88 (6): 304–16 '91

Contract, Social
- Freeman, Samuel: Contractualism, Moral Motivation, and Practical Reason. 88 (6): 281–303 '91

Decision Theory
- Godfrey-Smith, Peter: Signal, Decision, Action. 88 (12): 709–22 '91

Deontology
- Brook, Richard: Agency and Morality. 88 (4): 190–212 '91

Descriptions, Theory of
- Wilson, George M.: Reference and Pronominal Descriptions. 88 (7): 359–87 '91

Dilemma, Prisoner's
- Hollis, Martin: Penny Pinching and Backward Induction. 88 (9): 473–88 '91

Donnellan, Keith S.
- Wilson, George M : Reference and Pronominal Descriptions. 88 (7): 359–87 '91

Epistemology: *See* Knowledge, Theory of.

Equality
- Kittay, Eva Feder: In Whose Different Voice? [Abstract.] 88 (11): 645–6 '91
- Minow, Martha: Equalities. [Abstract.] 88 (11): 633–44 '91

Feminism

Hall, Pamela: Feminism and the Canon. [Abstract.] 88 (10). 568–9 '91

Kourany, Janet A.: Philosophy in a Different Voice. [Abstract.] 88 (10): 557–67 '91

Fichte, J. G.

Breazeale, Daniel: Why Fichte Now? [Abstract.] 88 (10): 524–31 '91

Jalloh, Chernor M.: Fichte: Foundationalism, Antifoundationalism, and the New Nihilism. [Abstract.] 88 (10): 542–3 '91

Game Theory

Hollis, Martin: Penny Pinching and Backward Induction. 88 (9): 473–88 '91

Geometry, Euclidean

Nerlich, Graham: How Euclidean Geometry Has Misled Metaphysics. 88 (4): 169–89 '91

Hegel, Georg Wilhelm Friedrich

Pippin, Robert B.: Idealism and Agency in Kant and Hegel. [Abstract.] 88 (10): 532–41 '91

Holism

Healey, Richard A.: Holism and Nonseparability. 88 (8): 393–421 '91

Human Action: *See* Action, Human.

Idealism

Pippin, Robert B.: Idealism and Agency in Kant and Hegel. [Abstract.] 88 (10): 532–41 '91

Identity

Ayers, Michael: Substance: Prolegomena to a Realist Theory of Identity. 88 (2): 69–90 '91

Identity Theory

Fodor, Jerry A.: A Modal Argument for Narrow Content. 88 (1): 5–26 '91

Indeterminacy

Gemes, Ken. The Indeterminacy Thesis Reformulated. 88 (2). 91–108 '91

Induction

Hollis, Martin: Penny Pinching and Backward Induction. 88 (9): 473–88 '91

Justification

Kaplan, Mark: Epistemology on Holiday. 88 (3): 132–54 '91

Kant, Immanuel

Pippin, Robert B.: Idealism and Agency in Kant and Hegel. [Abstract.] 88 (10): 532–41 '91

Knowledge, Theory of

Godfrey-Smith, Peter: Signal, Decision, Action 88 (12). 709–22 '91

Goldman, Alvin I.: Epistemic Paternalism: Communication Control in Law and Society. 88 (3): 113–31 '91

Hardwig, John: The Role of Trust in Knowledge. 88 (12): 693–708 '91

Jenkins, John: Aquinas on the Veracity of the Intellect. [Abstract.] 88 (11): 623–32 '91

Kaplan, Mark: Epistemology on Holiday 88 (3): 132–54 '91

Kitcher, Philip: Socializing Knowledge. [Abstract.] 88 (11): 675–6 '91

Laudan, Larry: Empirical Equivalence and Underdetermination. By Larry Laudan and Jarrett Leplin. 88 (9): 449–72 '91

Leplin, Jarrett: Empirical Equivalence and Underdetermination. By Larry Laudan and Jarrett Leplin. 88 (9): 449–72 '91

Longino, Helen E.: Multiplying Subjects and the Diffusion of Power. [Abstract.] 88 (11): 666–74 '91

Rouse, Joseph: The Dynamics of Power and Knowledge in Science. [Abstract.] 88 (11): 658–65 '91

Kripke, Saul

Wilson, George M.: Reference and Pronominal Descriptions. 88 (7): 359–87 '91

Language, Philosophy of

Tirrell, Lynne: Reductive and Nonreductive Simile Theories of Metaphor 88 (7): 337–58 '91

Law, Philosophy of

Goldman, Alvin I. Epistemic Paternalism: Communication Control in Law and Society. 88 (3): 113–31 '91

Kittay, Eva Feder: In Whose Different Voice? [Abstract.] 88 (11): 645–6 '91

Minow, Martha: Equalities. [Abstract.] 88 (11): 633–44 '91

Leibniz, Gottfried Wilhelm

Nerlich, Graham: How Euclidean Geometry Has Misled Metaphysics. 88 (4): 169–89 '91

Life, Human

Harries, Karsten: Questioning the Question of the Worth of Life. [Abstract.] 88 (11): 684–90 '91

McDermott, John J.: Why Bother: Is Life Worth Living? [Abstract.] 88 (11): 677–83 '91

Logic

Almog, Joseph: The Plenitude of Structures and Scarcity of Possibilities. [Abstract.] 88 (11): 620–2 '91

Almog, Joseph: The What and the How. 88 (5): 225–44 '91

Bricker, Phillip: Plenitude of Possible Structures. [Abstract.] 88 (11): 607–19 '91

Creath, Richard: Convention, Neutrality, and the Limits of Logic. [Abstract.] 88 (10): 522–3 '91

Sorensen, Roy A.: 'P, therefore, P' without Circularity. 88 (5): 245–66

Logical Positivism

Friedman, Michael: The Re-evaluation of Logical Positivism. [Abstract.] 88 (10): 505–19 '91

Weissman, David: Logical Positivism: A Retrospective. [Abstract.] 88 (10): 520–1 '91

Materialism

Dennett, Daniel C.: Real Patterns. 88 (1): 27–51 '91

Maudlin, Tim

Barnes, Eric. The Causal History of Computational Activity: Maudlin and Olympia. 88 (6): 304–16 '91

Metaphor

Tirrell, Lynne: Reductive and Nonreductive Simile Theories of Metaphor. 88 (7): 337–58 '91

Metaphysics

Healey, Richard A: Holism and Nonseparability. 88 (8): 393–421 '91

Nerlich, Graham: How Euclidean Geometry Has Misled Metaphysics. 88 (4): 169–89 '91

See also Ontology.

Mind, Philosophy of: *See* Identity Theory.

Motivation

Freeman, Samuel: Contractualism, Moral Motivation, and Practical Reason. 88 (6): 281–303 '91

Music

Alperson, Philip: The Arts of Music. [Abstract.] 88 (10): 555–6 '91

Kivy, Peter: Is Music an Art? [Abstract.] 88 (10): 544–54 '91

Naturalism

Bedau, Mark: Can Biological Teleology Be Naturalized? [Abstract.] 88 (11): 647–55 '91

Matthen, Mohan: Naturalism and Teleology. [Abstract.] 88 (11): 656–7 '91

Natural Kinds

Fodor, Jerry A.: A Modal Argument for Narrow Content. 88 (1): 5–26 '91

New Books.

88 (7): 388–92, 88 (9): 502–4

New Books: Anthologies.

88 (4): 223–4, 88 (8): 445–6

New Books. Translations.

88 (5). 280, 88 (8): 447–8

New Books: Paperbacks.

88 (8): 446–7

Notes and News.

88 (2): 112, 88 (3): 167–8, 88 (9). 501, 88 (11) 691–2

Ontology

Almog, Joseph: The What and the How. 88 (5): 225–44 '91
Healey, Richard A.: Holism and Nonseparability. 88 (8): 393–421 '91
See Also **Metaphysics.**

Possibility

Almog, Joseph: The Plenitude of Structures and Scarcity of Possibilities. [Abstract.] 88 (11): 620–2 '91
Bricker, Phillip: Plenitude of Possible Structures [Abstract.] 88 (11): 607–19 '91

Quantum Mechanics

Healey, Richard A.: Holism and Nonseparability. 88 (8): 393–421 '91

Quine, W. V.

Gemes, Ken: The Indeterminacy Thesis Reformulated. 88 (2): 91–108 '91

Realism

Ayers, Michael: Substance: Prolegomena to a Realist Theory of Identity. 88 (2): 69–90 '91
Dennett, Daniel C.: Real Patterns. 88 (1): 27–51 '91

Reason

Freeman, Samuel: Contractualism, Moral Motivation, and Practical Reason. 88 (6): 281–303 '91

Reference

Gemes, Ken: The Indeterminacy Thesis Reformulated. 88 (2): 91–108 '91

Scheffler, Samuel

Brook, Richard: Agency and Morality. 88 (4): 190–212 '91

Science, Philosophy of

Laudan, Larry: Empirical Equivalence and Underdetermination. By Larry Laudan and Jarrett Leplin. 88 (9): 449–72 '91
Leplin, Jarrett: Empirical Equivalence and Underdetermination. By Larry Laudan and Jarrett Leplin. 88 (9): 449–72 '91

Space

Nerlich, Graham: How Euclidean Geometry Has Misled Metaphysics. 88 (4): 169–89 '91

Substance

Ayers, Michael: Substance: Prolegomena to a Realist Theory of Identity. 88 (2): 69–90 '91

Teleology

Bedau, Mark: Can Biological Teleology Be Naturalized? [Abstract.] 88 (11): 647–55 '91
Matthen, Mohan: Naturalism and Teleology. [Abstract.] 88 (11): 656–7 '91

Trust

Hardwig, John: The Role of Trust in Knowledge. 88 (12): 693–708 '91

Truth, Theory of

Almog, Joseph: The What and the How. 88 (5): 225–44 '91

INDEX OF AUTHORS

Including Contributing Authors and Reviewers

Almog, Joseph, The Plenitude of Structures and Scarcity of Possibilities. [Abstract.] 88 (11): 620–22 '91

Almog, Joseph, The What and the How. 88 (5): 225–44 '91

Alperson, Philip, The Arts of Music. [Abstract.] 88 (10): 555–6 '91

Avramides, Anita, Review of Arthur W. Collins: *The Nature of Mental Things*. 88 (1): 52–6 '91

Ayers, Michael, Substance: Prolegomena to a Realist Theory of Identity. 88 (2): 69–90 '91

Barnes, Eric, The Causal History of Computational Activity: Maudlin and Olympia. 88 (6): 304–16 '91

Bedau, Mark, Can Biological Teleology Be Naturalized? [Abstract.] 88 (11): 647–55 '91

Breazeale, Daniel, Why Fichte Now? [Abstract.] 88 (10): 524–31 '91

Bricker, Phillip, Plenitude of Possible Structures. [Abstract.] 88 (11): 607–19 '91

Brook, Richard, Agency and Morality. 88 (4): 190–212 '91

Creath, Richard, Convention, Neutrality, and the Limits of Logic. [Abstract.] 88 (10): 522–3 '91

Dennett, Daniel C., Real Patterns. 88 (1): 27–51 '91

Fischer, John Martin, Review of William Hasker: *God, Time, and Knowledge* and Edward R Wierenga: *The Nature of God: An Inquiry into Divine Attributes*. 88 (8): 427–33 '91

Fodor, Jerry A., A Modal Argument for Narrow Content. 88 (1): 5–26 '91

Fogelin, Robert J., Review of Paul Grice: *Studies in the Way of Words*. 88 (4): 213–9 '91

Freeman, Samuel, Contractualism, Moral Motivation, and Practical Reason. 88 (6): 281–303 '91

Friedman, Michael, The Re-evaluation of Logical Positivism. [Abstract.] 88 (10): 505–19 '91

Gemes, Ken, The Indeterminacy Thesis Reformulated. 88 (2): 91–108 '91

Godfrey-Smith, Peter, Signal, Decision, Action. 88 (12): 709–22 '91

Goldman, Alvin I., Epistemic Paternalism: Communication Control in Law and Society. 88 (3): 113–31 '91

Hacking, Ian, Review of Paul Feyerabend: *Against Method* and *Farewell to Reason*. 88 (4): 219–23 '91

Hall, Pamela, Feminism and the Canon. [Abstract.] 88 (10): 568–9 '91

Hampton, Jean, Review of Jon Elster: *Cement of Society*. 88 (12): 728–38 '91

Hardwig, John, The Role of Trust in Knowledge 88 (12): 693–708 '91

Harries, Karsten, Questioning the Question of the Worth of Life. [Abstract.] 88 (11): 684–90 '91

Healey, Richard A., Holism and Nonseparability. 88 (8): 393–421 '91

Hill, Thomas E., Jr., Review of Rüdiger Bittner: *What Reason Demands*. 88 (9): 497–501 '91

Hollis, Martin, Penny Pinching and Backward Induction. 88 (9): 473–88 '91

Hughes, R. I. G., Review of Arthur Fine: *The Shaky Game: Einstein, Realism, and the Quantum Theory*. 88 (5): 275–9 '91

Hursthouse, Rosalind, Arational Actions. 88 (2): 57–68 '91

Jacob, Pierre, Review of Colin McGinn: *Mental Content*. 88 (12): 723–8 '91

Jalloh, Chernor M., Fichte: Foundationalism, Antifoundationalism, and the New Nihilism. [Abstract.] 88 (10): 542–3 '91

Jenkins, John, Aquinas on the Veracity of the Intellect. [Abstract.] 88 (11): 623–32 '91

Kaplan, Mark, Epistemology on Holiday. 88 (3): 132–54 '91

Kitcher, Philip, Review of Ronald N. Giere: *Explaining Science: A Cognitive Approach*. 88 (3): 163–7 '91

Kitcher, Philip, Socializing Knowledge. [Abstract.] 88 (11): 675–6 '91

Kittay, Eva Feder, In Whose Different Voice? [Abstract.] 88 (11): 645–6 '91

Kivy, Peter, Is Music an Art? [Abstract.] 88 (10): 544–54 '91

Kosso, Peter, Review of Richard W. Miller: *Fact and Method*. 88 (3): 159–62

Kourany, Janet A., Philosophy in a Different Voice. [Abstract.] 88 (10): 557–67 '91

Kyburg, Henry E., Jr., Review of Judea Pearl: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 88 (8): 434–7 '91

Laudan, Larry, Empirical Equivalence and Underdetermination. By Larry Laudan and Jarrett Leplin. 88 (9): 449–72 '91

Leplin, Jarrett, Empirical Equivalence and Underdetermination. By Larry Laudan and Jarrett Leplin. 88 (9): 449–72 '91

Levi, Isaac, Review of Peter Gärdenfors: *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. 88 (8): 437–44 '91

Levine, Andrew, Review of G. A. Cohen: *History, Labour and Freedom: Themes from Marx*. 88 (5): 267–75 '91

Longino, Helen E., Multiplying Subjects and the Diffusion of Power. [Abstract.] 88 (11): 666–74 '91

Luban, David, Review of Stuart Hampshire: *Innocence and Experience*. 88 (6): 317–24 '91

Ludlow, Peter, Review of Eva Kittay: *Metaphor: Its Cognitive Force and Linguistic Structure*. 88 (6): 324–30 '91

Matthen, Mohan, Naturalism and Teleology. [Abstract.] 88 (11): 656–7 '91

McDermott, John J., Why Bother: Is Life Worth Living? [Abstract.] 88 (11): 677–83 '91

Minow, Martha, Equalities. [Abstract.] 88 (11): 633–44 '91

Nerlich, Graham, How Euclidean Geometry Has Misled Metaphysics. 88 (4): 169–89 '91

Pippin, Robert B., Idealism and Agency in Kant and Hegel. [Abstract.] 88 (10): 532–41 '91

Rouse, Joseph, The Dynamics of Power and Knowledge in Science. [Abstract.] 88 (11): 658–65 '91

Ryan, Alan, Review of Jeremy Waldron: *The Right to Private Property*. 88 (3): 155–9 '91

Schneewind, J. B., Review of Charles Taylor: *Sources of the Self.* 88 (8): 422–6 '91

Sorensen, Roy A., 'P, therefore, P' without Circularity. 88 (5): 245–66

Steiner, Mark, Review of Michael Detlefsen: *Hilbert's Program: An Essay on Mathematical Instrumentalism.* 88 (6): 331–6 '91

Striker, Gisela, Review of T. H. Irwin: *Aristotle's First Principles.* 88 (9). 489–96 '91

Tirrell, Lynne, Reductive and Nonreductive Simile Theories of Metaphor. 88 (7): 337–58 '91

Weissman, David, Logical Positivism: A Retrospective. [Abstract.] 88 (10): 520–1 '91

Wilson, George M., Reference and Pronominal Descriptions. 88 (7): 359–87 '91

Yolton, John, Review of Steven M. Nadler: *Arnauld and the Cartesian Philosophy of Ideas.* 88 (2): 109–12 '91

THE JOURNAL OF PHILOSOPHY

SUBSCRIPTIONS (12 issues)

Individuals	$30.00
Libraries and Institutions	$50.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$ 8.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are taken on a first come, first served basis.

Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, PA. Typeset by TAPSCO, Inc., Akron, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, 709 Philosophy Hall, Columbia University, New York, New York, 10027. (212) 866-1742

INDEX OF BOOKS REVIEWED

Listed by Author of Item Reviewed

Bittner, Rüdiger: *What Reason Demands.*—Thomas E. Hill, Jr. 88 (9): 497–501 '91

Cohen, G. A.: *History, Labour and Freedom: Themes from Marx.*—Andrew Levine. 88 (5): 267–75 '91

Collins, Arthur W.: *The Nature of Mental Things.*—Anita Avramides. 88 (1): 52–6 '91

Detlefsen, Michael: *Hilbert's Program: An Essay on Mathematical Instrumentalism.*—Mark Steiner. 88 (6): 331–6 '91

Elster, Jon: *Cement of Society.*—Jean Hampton. 88 (12): 728–38 '91

Feyerabend, Paul: *Against Method* and *Farewell to Reason.*—Ian Hacking. 88 (4): 219–23 '91

Fine, Arthur: *The Shaky Game: Einstein, Realism, and the Quantum Theory.*—R. I G. Hughes. 88 (5): 275–9 '91

Gärdenfors, Peter: *Knowledge in Flux: Modeling the Dynamics of Epistemic States* —Isaac Levi. 88 (8): 437–44 '91

Giere, Ronald N.: *Explaining Science: A Cognitive Approach.*—Philip Kitcher. 88 (3): 163–7 '91

Grice, Paul: *Studies in the Way of Words.*—Robert J. Fogelin. 88 (4): 213–9 '91

Hampshire, Stuart: *Innocence and Experience.*—David Luban. 88 (6): 317–24 '91

Hasker, William: *God, Time, and Knowledge.*—John Martin Fischer. 88 (8): 427–33 '91

Irwin, T. H.: *Aristotle's First Principles.*—Gisela Striker. 88 (9): 489–96 '91

Kittay, Eva: *Metaphor: Its Cognitive Force and Linguistic Structure.*—Peter Ludlow. 88 (6) 324–30 '91

McGinn, Colin: *Mental Content.*—Pierre Jacob. 88 (12): 723–38 '91

Miller, Richard W.: *Fact and Method.*—Peter Kosso. 88 (3). 159–62 '91

Nadler, Steven M.: *Arnauld and the Cartesian Philosophy of Ideas.*—John Yolton. 88 (2): 109–12 '91

Pearl, Judea: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.*—Henry E. Kyburg, Jr. 88 (8): 434–7 '91

Taylor, Charles: *Sources of the Self.*—J. B Schneewind. 88 (8): 422–6 '91

Waldron, Jeremy: *The Right to Private Property.*—Alan Ryan. 88 (3): 155–9 '91

Wierenga, Edward R.: *The Nature of God: An Inquiry into Divine Attributes.*—John Martin Fischer. 88 (8): 427–33 '91

THE JOURNAL OF PHILOSOPHY

is pleased to announce the publication of its

TWENTY-FIVE YEAR INDEX, 1964–1988

VOLUME LXI TO VOLUME LXXXV
AUTHORS, SUBJECTS, & REVIEWS

Supplement to its

FIFTY-YEAR INDEX, 1904–1953

TEN-YEAR SUPPLEMENT, 1954–1963

It is now available at special prices:

25-year (1964–1988) $25.00 cloth/$10 00 paper

also available

50-year (1904–1953) $12.00 cloth 10-year (1954–1963) $3.50 paper

Shipping $5.00 (first)/$2 50 (each addl.)

Please send a check or money order to:

Journal of Philosophy
704 Philosophy Hall
Columbia University
New York, New York 10027

Prices are in effect through May 31, 1992 Orders and payment must be received by then Paper copies will be mailed immediately, but cloth copies will take 6–8 weeks for delivery

---

Statement of Ownership, Management and Circulation

(Required by 39 U S C 3685)

1A Title of Publication
THE JOURNAL OF PHILOSOPHY

1B PUBLICATION NO.
0 0 2 2 3 6 2 X

2 Date of Filing
9/9/91

3 Frequency of Issue
MONTHLY

3A No. of Issues Published Annually

3B Annual Subscription Price

4 Complete Mailing Address of Known Office of Publication (Street, City, County, State and ZIP+4 Code) (Not printers)
709 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NEW YORK NEW YORK 10027

5 Complete Mailing Address of the Headquarters of General Business Offices of the Publisher (Not printer)
709 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NEW YORK NEW YORK 10027

6 Full Names and Complete Mailing Address of Publisher, Editor, and Managing Editor (This item MUST NOT be blank)

Publisher (Name and Complete Mailing Address)
THE JOURNAL OF PHILOSOPHY, 709 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NEW YORK NEW YORK 10027

Editor (Name and Complete Mailing Address)
ARTHUR DANTO, 709 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NEW YORK NY 10027

Managing Editor (Name and Complete Mailing Address)
MICHAEL KELLY, 704 PHILOSOPHY HALL, COLUMBIA UNIVERSITY NEW YORK NY 10027

7 Owner [illegible]

Full Name	Complete Mailing Address
THE JOURNAL OF PHILOSOPHY	709 PHILOSOPHY HALL COLUMBIA UNIVERSITY NEW YORK NEW YORK 10027

8 Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages or Other Securities (If there are none, so state)

Full Name	Complete Mailing Address
NONE	

9 For Completion by Nonprofit Organizations Authorized To Mail at Special Rates (DMM Section 424.12 only)
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes (Check one)

(1) ☒ Has Not Changed During Preceding 12 Months

(2) ☐ Has Changed During Preceding 12 Months

(If changed, publisher must submit explanation of change with this statement.)

10 Extent and Nature of Circulation (See instructions on reverse side)	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
A Total No. Copies (Net Press Run)	4845	4743
B Paid and/or Requested Circulation 1 Sales through dealers and carriers, street vendors and counter sales	——	——
2 Mail Subscription (Paid and/or requested)	3972	3915
C Total Paid and/or Requested Circulation (Sum of 10B1 and 10B2)	3972	3915
D Free Distribution by Mail, Carrier or Other Means Samples, Complimentary, and Other Free Copies	174	178
E Total Distribution (Sum of C and D)	4146	4093
F Copies Not Distributed 1 Office use, left over, unaccounted, spoiled after printing	710	651
2 Return from News Agents	——	——
G TOTAL (Sum of E, F1 and 2—should equal net press run shown in A)	4856	4744

11 I certify that the statements made by me above are correct and complete

Signature and Title of Editor, Publisher, Business Manager, or Owner
Michael Kelly, Managing Editor

PS Form 3526, January 1991 (See instructions on reverse)

STUDIES IN CONTINENTAL THOUGHT

The Principle of Reason

Martin Heidegger
Translated by Reginald Lilly

In this text of a lecture course that he gave in 1955-56, Martin Heidegger presents his most extensive reflection on the notion of history and its essence, the Geschick of being.

cloth $27.50

Narrative and the Self

Anthony Paul Kerby

Invoking the writings of Benveniste, Ricoeur, Merleau-Ponty, Lacan, Taylor, and other theorists, Kerby argues that language and narration play a central role in key aspects of human experience such as emotion, values, recollection, and sense of history.

cloth $22.50

Sartre's Political Theory

William L. McBride

"... the most perceptive analysis of Sartre's theory of history that I have read" —Thomas R. Flynn

McBride's detailed study of Sartre's political philosophy takes his twin ideals of "Socialism and Freedom" as a guiding theme.

cloth $37.50 paper $12.95

Writings of Charles S. Peirce

A Chronological Edition
Volume 5, 1884-1886

Compiled by the Editors of the Peirce Edition Project

"... an important event for the world of philosophy. For the first time we have available in an intelligible form the writings of one of the greatest philosophers of the past hundred years." —*The Times Literary Supplement*

Volume 5 of this landmark edition covers an important transition in Peirce's life, marked by a rekindled enthusiasm for speculative philosophy.

cloth $65.00

INDIANA UNIVERSITY PRESS

Feminism and Philosophy

Perspectives on Difference and Equality

[illegible]

"[illegible] ... A very [illegible] text." —[illegible]

cloth [illegible] paper [illegible]

Fragments of Redemption

Jewish Thought and Literary Theory in Benjamin, Scholem, and Levinas

[illegible]

[illegible]

cloth [illegible] paper [illegible]

Kant's Philosophy of Religion Reconsidered

Edited by Philip J. Rossi and Michael J. Wreen

Challenging the traditional view that Kant's account of religion was peripheral to his thinking, these essays demonstrate the centrality of religion to Kant's critical philosophy

The Indiana Series in the Philosophy of Religion

cloth $25.00

Princeton

New in paperback

The Concept of Irony

With Continual Reference to Socrates Together with Notes of Schelling's Berlin Lectures

Søren Kierkegaard

Edited and translated, with introduction and notes, by Howard V. Hong and Edna H. Hong

A work that "not only treats of irony but is irony," wrote a contemporary reviewer of *The Concept of Irony, with Continual Reference to Socrates.* Presented here with Kierkegaard's notes of the celebrated Berlin lectures on "positive philosophy" by F. W. J. Schelling, the book is a seedbed of Kierkegaard's subsequent work, both stylistically and thematically.

Now in paper: $17.95 ISBN 0-691-02072-8

Practice in Christianity

Søren Kierkegaard

Edited and translated, with introduction and notes, by Howard V. Hong and Edna H. Hong

Of the many works he wrote during 1848, his "richest and most fruitful year," Kierkegaard specified *Practice in Christianity* as "the most perfect and truest thing."

In his reflections on such topics as Christ's invitation to the burdened, the *imitatio Christi*, the possibility of offense, and the exalted Christ, he takes as his theme the requirement of Christian ideality in the context of divine grace. Kierkegaard asserts the need for institutional and personal admission of the accommodation of Christianity to the culture and to the individual misuse of grace.

Paper: $14.95 ISBN 0-691-02063-9 Cloth: $45.50 ISBN 0-691-07396-1

Princeton University Press

41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 258-4900
ORDERS: 800-PRS-ISBN (777-4726) • OR FROM YOUR LOCAL BOOKSTORE

NEW FROM MIT/BRADFORD BOOKS

Representation and Mind Series

The Realistic Spirit
Wittgenstein, Philosophy, and the Mind
Cora Diamond

"Diamond, more than any other commentator, conveys an accurate sense of Wittgenstein's methods and goals in philosophy Her essays illuminating the complicated subtle relationships between the earlier and later works are the best in the field "— *Thomas Ricketts, University of Pennsylvania*

A Bradford Book 408 pp $32 50

The Unity of the Self
Stephen L. *White*

A Bradford Book 488 pp $35 00

The Imagery Debate
Michael Tye

Michael Tye untangles the complex web of empirical and conceptual issues of the imagery debate in psychology

A Bradford Book 200 pp $22 50

New in paperback

Representation and Reality
Hilary Putnam

A Bradford Book 154 pp $9 95 paper

Learnability and Cognition
The Acquisition of Argument Structure
Steven Pinker

Caring About Morality
Philosophical Perspectives in Moral Psychology
Thomas E *Wren*

"I find Wren the thinker and writer who has best combined the various threads of motivational theory from psychology and philosophy. He offers to philosophers and others an unrivaled sensitivity to the inside story in psychology "— *John Broughton, Teachers College, Columbia University*

A Bradford Book 208 pp $25 00

The Incomplete Universe
Totality, Knowledge, and Truth
Patrick Grim

Grim examines a series of logical paradoxes to reveal their implications for contemporary epistemology, metaphysics, and the philosophy of religion

A Bradford Book 256 pp $25 00

The Bounds of Logic
A Generalized Viewpoint
Gila Sher

A Bradford Book 160 pp $27 50

Events in the Semantics of English
Terence Parsons

Current Studies in Linguistics No 19
334 pp $35 00

To order call toll-free 1-800-356-0343 Fax (617) 625-6660 MasterCard & VISA accepted

THE MIT PRESS
55 Hayward Street, Cambridge, MA 02142

NEW FROM THE FREE PRESS

New Third Edition —
Now Available for Immediate Adoption!

GREEK PHILOSOPHY

Thales to Aristotle

Third Edition, Revised and Expanded

Reginald E. Allen, Editor

Widely praised for its concentration on the issues that are most central to the history of Greek philosophy as well as for its accessibility, *Greek Philosophy: Thales to Aristotle* offers a valuable introduction to the works of the Presocratics, Plato, and Aristotle. Emphasizing metaphysics and epistemology as the foci of Greek philosophy, Allen's book also affords a solid basis for studying the majority of post-classical philosophers who have further explored metaphysics and epistemology as their central concerns.

Including key texts in Socratic ethics, Platonic politics, and natural science, the complete scope of Allen's collection has made it a perennial favorite, the most popular reader available on Greek philosophy. While the third edition retains the earlier emphasis on metaphysics and epistemology, it is enlarged to include more material on Greek science and mathematics, particularly with regards to analysis of infinity. Some of Aristotle's less well-known texts from his Physics have also been added. To complete the picture, revised translations of Plato's Symposium and Aristotle's Categories render these texts more faithful to the original Greek.

Reginald E. Allen is professor of classics and philosophy at Northwestern University.

1991 ISBN: 0-02-900495-0 $14.95 paper

Attention College Professors:
Order your examination copy today! Simply place your request on university stationery, indicating course name, current text, projected enrollment, and the date your next text decision is due.
Send to

THE FREE PRESS
A Division of Macmillan/A Maxwell Macmillan Company
Attn: Dino Battista,
866 Third Avenue, New York, NY 10022

Why Me?

A Philosophical Inquiry into Fate

Michael Gelven

"*Why Me?* is a profound philosophical reflection on the notion of fate, and to read this book is to join in with a thoughtful and strong-minded philosopher thinking through some extraordinarily important topics."
—Thomas V. Morris

In this stimulating inquiry into the nature of our existence, Gelven seeks to answer the question "what is fate?" in terms of what it means to be fated. Using four models—the gambler, the birthday celebrant, the historian, and the tragedian—corresponding to chance, destiny, fortune, and inevitability, Gelven provides a sophisticated yet lively discussion that ultimately relates fate to the great human issues of freedom, mortality, rationality, and truth.
209 pages $22.00 cloth

A Commentary on Heidegger's *Being and Time* Revised Edition

Michael Gelven

"Gelven is well prepared to go the arduous path of *Being and Time* from beginning to end commenting, criticising, elaborating chapter by chapter in a refreshingly lively manner." —*History of Philosophy*

Gelven "does . . . for Heidegger what Heidegger did for Hegel."
—*Journal of the History of Philosophy*
244 pages $30.00 cloth $9.50 paper

Quest for the Absolute

The Philosophical Vision of Joseph Maréchal

Anthony M. Matteo

This first full-length study of Maréchal to appear in English provides a clear reading of the complex thought of an important twentieth-century epistemological realist. Focusing on *Le point de depart de la metaphysique,* Matteo shows how Maréchal laid the groundwork for the integration of the Aristotelian-Thomistic ontological approach to knowledge and the transcendental analysis inaugurated by Kant. In his discussion, Matteo argues that Maréchal is especially relevant to current debates centering on the foundations of rationality and the renewed controversy surrounding cognitive relativism.
Approx. 250 pages $30.00 cloth March

Northern Illinois University Press DeKalb, Illinois 60115 (815) 753-1075

The Journal of Philosophy

Missing Issues?

Complete volumes and all separate issues are available back to Vol. I, No. 1 (January 7, 1904). Prices, not including shipping charges, are as follows:

Single issues:

1904–1976	$3.00
1977–present	$4.00
Dewey Lectures	$5.00

Double issues, devoted to the symposia of the Annual Meetings of the American Philosophical Association, Eastern Division:

1943–1976	$7.00 each set
1977–present	$9.00 each set

Complete volumes:

1904–last year	$60.00 unbound

Indexes:

50-year (1904–1953)	$12.00 cloth
10-year (1954–1963)	$3.50 paper
25-year (1964–1988)	$10.00 paper

Please inquire for shipping charges and discounts on back issues.

709 Philosophy Hall, Columbia University
New York, New York 10027

Cambridge University Press

Consciousness and Revolution in Soviet Philosophy

From the Bolsheviks to Evald Ilyenkov

David Bakhurst

Focusing on the work of Evald Ilyenkov (1924–79), this first critical history of the philosophical culture of the USSR offers a unique insight into the world of Soviet philosophy, the place of politics within it, and its prospects in the age of ***glasnost*** and ***perestroika.***

Modern European Philosophy

$49 50 Hardcover / $17 95 Paper

The Semantic Tradition from Kant to Carnap

To the Vienna Station

J. Alberto Coffa

Linda Wessels, Editor

"This is an extraordinary and long-awaited book. Its history is meticulous, but it treats its subjects as philosophers to be responded to and even fought with, rather than as museum pieces It is worth the wait."

—*Richard Creath,*
Arizona State University

$49.50 Hardcover

Psychology

Stephen Everson, Editor

Examination of the theories of the ancient philosophers, the Presocratics and the Hellenistics through Plato and Plotinus, reveals that psychology had become an established discipline long before Descartes.

Contributors. *Julia Annas, Eyjólfur K Emilsson, Stephen Everson, Christopher Gill, R J Hankinson, T. H. Irwin, A A Long, Sabina Lovibond, Malcolm Schofield*

Companions to Ancient Thought 2

$54 50 Hardcover / $16 95 Paper

Understanding Action

An Essay on Reasons

Frederic Schick

This is an important new book about human motivation, about the reasons people have for their actions. It focuses on how people see or understand their situations, options, and prospects and by taking account of these understandings, Professor Schick is able to expand the current theory of decision and action.

$39 50 Hardcover / $11 95 Paper

Mind and Common Sense

Philosophical Essays on Common Sense Psychology

Radu J. Bogdan, Editor

These essays examine current controversies about the importance of common sense psychology for our understanding of the human mind. A variety of positions on common sense psychology from critical to supportive, from exegetical to speculative, are represented.

Contributors: *Jonathan Bennett, Paul Churchland, Robert Cummins, Colin McGinn, Adam Morton, Alexander Rosenberg, Kathleen V. Wilkes, Radu J Bogdan.*

$44 50 Hardcover

Automatism, Insanity, and the Psychology of Criminal Responsibility

A Philosophical Inquiry

Robert F. Schopp

This important new book provides the reader with an understanding of the significance of psychological states and processes for the attribution of criminal responsibility in a manner that is logically coherent, morally defensible, and consistent with research in psychopathology.

Cambridge Studies in Philosophy and Law

$39 50 Hardcover

Available in bookstores or write:

CAMBRIDGE UNIVERSITY PRESS

40 West 20th Street, New York, NY 10011-4211. Call toll-free 800-872-7423.

MasterCard/VISA accepted. Prices subject to change.